

Addis Ababa  
University  
(Since 1950)



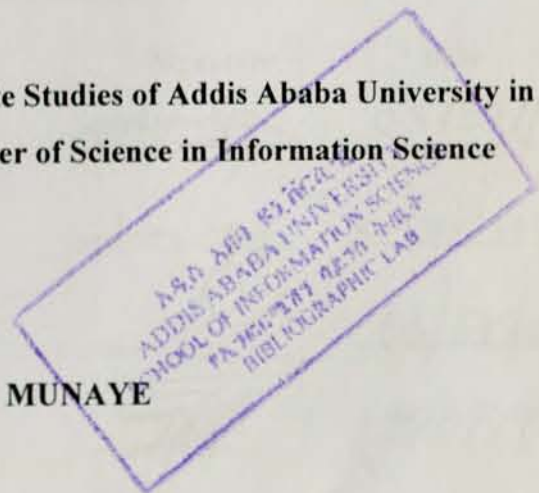
ADDIS ABABA UNIVERSITY  
SCHOOL OF GRADUATE STUDIES  
SCHOOL OF INFORMATION SCIENCE

DEVELOPING CASE BASED RECOMMENDER SYSTEM IN MICRO AND  
SMALL ENTERPRISES DEVELOPMENT AGENCY FOR THE SELECTION  
OF ENTERPRISES TO NEW MEMBERS: THE CASE OF ETHIOPIA

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in  
Partial Fulfillment for the Degree of Master of Science in Information Science

BY

YIRGA YAYEH MUNAYE



JUNE, 2014

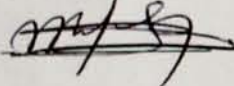
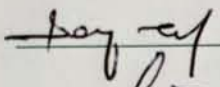
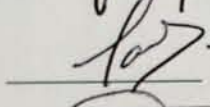
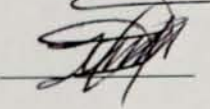
ADDIS ABABA UNIVERSITY  
SCHOOL OF GRADUATE STUDIES  
SCHOOL OF INFORMATION SCIENCE

DEVELOPING CASE BASED RECOMMENDER SYSTEM IN MICRO AND  
SMALL ENTERPRISES DEVELOPMENT AGENCY FOR THE SELECTION  
OF ENTERPRISES TO NEW MEMBERS: THE CASE OF ETHIOPIA

BY

YIRGA YAYEH

Names and Signature of Members of the Examining Board

Name	Signature	Date
<u>Miftah Hassen</u> Chair person, Examining Board		<u>08/07/14</u>
<u>Dereje Tefai (PhD)</u> Advisor		<u>08/07/2014</u>
<u>Rahel Bekele (PhD)</u> Examiner		<u>08/07/2014</u>
<u>Solomon Tegeha</u> Examiner		<u>08/07/14</u>

## DEDICATION

I Would Like to Dedicate this Thesis to My Father and Mother.

## ACKNOWLEDGEMENT

First of all, I would like to forward my genuine gratitude to my research advisor, Dr. Dereje Teferi, for his crucial guidance, encouragement and constructive comments about all parts of this research work.

I would like to thank many people for their kindness and assistance during my thesis work. I truly appreciate the support of the experts and all members at the Federal Micro and Small Enterprise Development Agency of Ethiopia.

My heartfelt thanks also go to my instructor, Dr. Million Meshesha, who provides unreserved precious advices, comments, and important directions on the research proposal development.

I wish to thank my beloved friends and classmates, especially for: Tamie, Haylie and Yibeltal who have one way or another contributed to the success of my thesis.

I truly appreciate my fiancée Bethelhem Yinesu and her mother Genet S., for their patient, honesty, support and advice with ups and downs of my life activities and this research work. I would like to say thank you to her friends to their support during this thesis work.

Above all, I would like to thank the Almighty God and who blessed each and every minute of my life in doing all my work in his way.

**Yirga Yayeh Munaye**

**June, 2014**

## TABLE OF CONTENTS

DEDICATION.....	I
ACKNOWLEDGEMENT.....	II
LIST OF FIGURES.....	viii
LIST OF TABLES.....	ix
LIST OF ACRONYMS.....	x
ABSTRACT.....	xi
CHAPTER ONE.....	1
1. INTRODUCTION.....	1
1.1. Background.....	1
1.2. Statement of the Problem.....	4
1.3. Objective of the Study.....	6
1.3.1. General Objective.....	6
1.3.2. Specific Objectives.....	6
1.4. Methodology.....	7
1.4.1. Literature Review.....	7
1.4.2. Method of Data Collection.....	7
1.4.3. Sampling Technique and Sample Size.....	7
1.4.4. Case Representation.....	8
1.4.5. Implementation Tool.....	9
1.4.6. Testing Procedures.....	9
1.5. Scope and Limitation of the Study.....	10
1.6. Significance of the Study.....	10
1.7. Organization of the Thesis.....	11

CHAPTER TWO.....	12
2. LITERATURE REVIEW.....	12
2.1. Recommender System.....	12
2.2. Architecture of Recommender system.....	13
2.3. Types of Recommender System.....	14
2.4. Knowledge based recommender systems.....	14
2.5. Knowledge based system (case based reasoning system).....	15
2.5.1. Architecture of knowledge based system.....	15
2.5.2. The Knowledge Development/Engineering process.....	17
2.5.2.1. Knowledge acquisition.....	17
2.5.2.2. Knowledge representation.....	20
2.5.2.3. Knowledge validation.....	20
2.6. Knowledge based reasoning techniques.....	20
2.6.1. Case based Reasoning.....	20
2.6.1.1. Case based reasoning cycle.....	22
2.7. CBR Techniques.....	26
2.7.1. Case Representation.....	26
2.7.2. Indexing.....	27
2.7.3. Storage.....	28
2.7.4. Retrieval.....	29
2.8. Case Adaptation.....	29
2.9. Advantage of case based reasoning.....	29
2.10. Disadvantage of case based reasoning.....	30
2.11. Integrating Rule-based and Case-based Reasoning.....	30

2.12. CBR System Performance Evaluation Methods.....	31
2.13. Knowledge based system development tools.....	31
2.14. MSES in Ethiopia.....	33
2.15. Opportunity Areas for MSEs in Ethiopia.....	36
2.16. Related Research works.....	36
CHAPTER THREE.....	40
3. KNOWLEDGE ACQUISITION AND MODELING.....	40
3.1. Knowledge Acquisition.....	40
3.1.1. Knowledge Acquisition from domain experts.....	40
3.1.2. Knowledge acquisition from Members.....	41
3.1.3. Knowledge Acquisition from Relevant Document.....	42
3.2. Attribute selection using data mining tools.....	42
3.2.1. Data Collection.....	42
3.2.2. Data Preparation and Cleaning.....	43
3.3. Knowledge Modeling.....	43
3.4. Enterprise sector and Enterprise activity selection.....	44
3.4.1. Factors to select Micro and Small enterprise sectors and activities.....	45
3.4.1.1. Age.....	45
3.4.1.2. Gender.....	47
3.4.1.3. Capital.....	48
3.4.1.4. Enterprise Location.....	49
3.4.1.5. Number of members.....	54
3.4.1.6. Educational level.....	54
3.4.1.7. Form of ownership.....	56

3.4.1.8. Enterprise type .....	56
3.5. Knowledge Representation .....	60
3.6. Case structure for enterprise sector and enterprise activity selection .....	61
CHAPTER FOUR .....	64
4. DESIGN AND IMPLEMENTATION OF THE PROTOTYPE .....	64
4.1. Designing the architecture of case based recommender system .....	64
4.2. Case based recommender system for the selection of enterprise sectors and activities .....	66
4.2.1. Building the Case Base .....	68
4.2.2. Case Representation .....	68
4.2.3. Case attributes description for CBRSESAEAS .....	69
4.2.4. Managing connectors .....	72
4.2.5. Managing Tasks .....	73
4.2.6. Managing methods .....	76
4.2.7. Deploy the case base recommender system .....	78
4.2.8. Explanation Facilities .....	79
CHAPTER FIVE .....	81
5. TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE .....	81
5.1. INTRODUCTION .....	81
5.2. Evaluation of the retrieval process .....	81
5.3. Evaluation of the reuse process .....	84
5.4. Comparison of the Performance of CBRSESAEAS with Previous CBR Systems .....	84
5.5. User Acceptance Testing .....	85
CHAPTER SIX .....	88
6. CONCLUSION AND RECOMMENDATIONS .....	88

6.1. Conclusion .....	88
6.2. Recommendation .....	89
REFERENCES .....	91
APPENDIX.....	95

## LIST OF FIGURES

Figure 2.1: Architecture of recommender system .....	13
Figure 2.2: Architecture of a Knowledge-Based System .....	16
Figure 2.3: Development of a Knowledge-Based System.....	17
Figure 2.4: Two Major Components of the CBR System .....	21
Figure 2.5: The case based reasoning cycle .....	22
Figure 2.6: Software Architecture of jCOLIBRI.....	33
Figure 3.1. Hierarchical structure of enterprise sector selection .....	44
Figure 3.2: The age factors in different category of age young, old or plc .....	47
Figure 3.3: The gender factors in different category of either male or female.....	48
Figure 3.4: The amount of capital required for starting an enterprise sectors .....	49
Figure 3.5: The location factor for the selection of enterprise sectors .....	50
Figure 3.6: Educational level factors for the development of enterprise.....	56
Figure 3.7: enterprise type factor for enterprise selection .....	57
Figure 4.1: Architecture of CBRSESAEAS .....	65
Figure 4.2: The Main Window of jCOLIBRI.....	66
Figure 4.3: New CBR application window.....	67
Figure 4.4: Types of jCOLIBRI extensions.....	68
Figure 4.5: Defining case structure.....	70
Figure 4.6: jCOLIBRI case base schemas .....	72
Figure 4.7: Managing connectors .....	73
Figure 4.8: Configure CBR application.....	74
Figure 4.9: Revision and retain tasks.....	76
Figure 4.10: Configuration of tasks and methods.....	78
Figure 4.11: Query entry window in to the case base.....	79
Figure 4.12: Explanation facility for enterprise activity.....	80

## LIST OF TABLES

Table 2.1: Previous related works .....	39
Table 3.1: Regions and enterprise sectors .....	54
Table 3.2: The Case Structure for enterprise sector selection .....	62
Table 4.1: Descriptions and Weight of the Selected Attributes.....	71
Table 5.1: relevant cases assigned by domain experts for the sample test case .....	82
Table 5.2: recall and precision results for the sample test case .....	83
Table 5.3: Accuracy value of the reuse process .....	84
Table 5.4: A comparison of CBRSESAEAS system with the previous CBR systems .....	85
Table 5.5: The CBRSESAEAS system performance evaluation by the domain expert.....	86
Table 5.6: The CBRSESAEAS system performance evaluation by the enterprise members .....	87

## LIST OF ACRONYMS

AI	-----	Artificial Intelligence
CBR	-----	Case Based Reasoning.
CBRSESAEAS	-----	Case Based Reasoning system for Enterprise Sector and Enterprise Activity Selection
CSV	-----	Comma Separated Values
ETB	-----	Ethiopian Birr
FeMSEDA	-----	Federal Micro and Small Enterprises Development Agency
GAIA	-----	Group of Artificial Intelligence Applications
GUI	-----	Graphical User interface
IE	-----	Inference Engine
JCOLIBRI	-----	java Cases and Ontology Libraries Integration for Building Reasoning Infrastructure
KBS	-----	Knowledge-based system
MSEs	-----	Micro and Small Enterprises
ML	-----	machine learning
Plc	-----	Private limited company

## ABSTRACT

Micro and small enterprise sectors are the most focused sectors especially in Ethiopia because they have a capacity to reduce unemployment, create job and generate high income. In this study, case based reasoning approach for the selection of Enterprise sector and activities is considered. In order to conduct the research, the required knowledge was collected from previous members' cases, domain experts, and documents through interview and document analysis methods. After all, the knowledge is modeled in hierarchical tree model and case based reasoning system is constructed.

The prototype of case based reasoning for enterprise sector and activity selection is implemented by jCOLIBRI 1.1 programming tool. Nearest neighbor retrieval algorithm is used to measure the similarity of new case (query) with cases in the case base. As a result, if there is a similarity between the new case and the existing case the system assigns the solution of previous case as a solution to new case. To determine the applicability of the prototype system in the domain area, the system has been evaluated by domain experts and members through visual interaction based on the criteria of ease of use, efficiency, applicability in the domain area and providing correct recommendation.

Based on the evaluation through user acceptance 87% system performance is obtained from domain experts and 86.5% system performance is obtained from enterprise members. And also the performance of the prototype is measured using recall, precision and accuracy measures, where the system achieves 86% recall, 64% precision and 90% accuracy.

Finally, conclusions and future research directions are forwarded.

## CHAPTER ONE

### 1. INTRODUCTION

#### 1.1. Background

Recommender systems were originally defined as systems in which “people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients”. The term now has a broader connotation, describing any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options. Such systems have an obvious appeal in an environment where the amount of on-line information vastly outstrips any individual’s capability to survey it [6].

Recommender systems provide advice to users about items they might wish to purchase or examine. Recommendations made by such systems can help users navigate through large information spaces of product descriptions, news articles or other items. Specifically, recommender systems have (i) background data, the information that the system has before the recommendation process begins, (ii) input data, the information that user must communicate to the system in order to generate a recommendation, and (iii) an algorithm that combines background and input data to arrive at its suggestions [6], [8].

Recommender systems may be based on collaborative filtering (by user ratings), These systems aggregate data about customers’ purchasing habits or preferences, and make recommendations to other users based on similarity in overall purchasing patterns. Content-based filtering (by keywords), are classifier systems derived from machine learning research. knowledge based recommender system that uses knowledge about users and products to pursue a knowledge based approach to generating a recommendation by reasoning about what products meet the user’s requirements and hybrid filtering (by both collaborative and content-based filtering) [6], [7].

Case-Based Reasoning (CBR) is one of the most successful machine learning methodologies that exploit a knowledge-rich representation of the application domain. Basically, CBR is a problem

solving methodology that addresses a new problem by first retrieving a past, already solved similar case, and then reusing that case for solving the current problem [11].

A case-based reasoning system solves new problems by retrieving old problems likely to have similar solutions. Researchers working on the retrieval of CBR cases have concentrated on developing knowledge-based methods for precise, efficient retrieval of well-represented examples [3].

A case based recommender system maintains a set of cases of previously solved recommendation problems and their solutions. According to [7], case base is the product catalogue which is the solutions of the recommendation problem. The problem is the user's query that is essentially a partial description of the desired product. A case could be thought of as a record in a database; collection of features and their values.

The effectiveness of the recommendation in case base recommender system is based on the ability to match user preferences with product description. The tools used to explain the match and to enforce the validity of the suggestion, the range of available functionalities and the graphical interface that support the user in browsing the recommended information content [8].

Recommendation process is a sequential process. It is not a simply sequential prediction problem rather a sequential decision problem. At each point, the recommender system makes a decision about the item/items. This decision should take into account the sequential process involved and the optimization criteria suitable for the recommender system. They suggest items to the user who can then accept one of the recommendations. At the next stage a new list of recommended items is calculated and presented to the user. This sequential nature of the recommendation process, where at each stage a new list is calculated based on the user's past ratings, will lead us naturally to our reformulation of the recommendation process as a sequential optimization process [9].

The Ethiopian government recognizes the significance of MSEs (Micro and Small Enterprises) sector and shows its dedication to promote the MSE Development by the Issuance of National MSEs Strategy in 1997 and the Establishment of the Federal Micro and Small Enterprises Development Agency (FeMSEDA). Ethiopia's industrial development strategy issued in 2003

also singled out the promotion of MSEs development as one of the important instruments to create productive and dynamic private sector. The promotion of this sector is justified on the grounds that enhancing growth with equity, creating long-term jobs, providing the basis for medium and large enterprise and promoting exports etc. The strategy puts a means to support the MSEs such as, infrastructure, financial facilities, supply of raw materials, and training [2].

The MSEs are also described as the national home of entrepreneurship. It provides the ideal environment enabling entrepreneurs to exercise their talents, to experiment, to innovate and transform their ideas into goods and services. In all the successful economies, MSEs are seen as an essential springboard for growth, job creation and social progress at large. The small business sector is also seen as an important force to generate employment and more equitable income distribution, to activate competition, exploit niche markets, and enhance productivity and technical change. The sector is characterized by low enterprise requirement, operational flexibility, and location wise mobility [4], [18].

The role of MSEs sector is characterized by a range of highly diversified activities, which can create job opportunities for a substantial segment of the population. This suggests that the sector may be a quick remedy for the unemployment problem and facilitate a suitable environment for new job seekers and for self-employment, particularly for the urban population [1], [2], [4].

According to the manual of FeMSEDA (2013) the specific MSEs opportunities in Ethiopia promoted by the state are in Manufacturing (wood work including furniture, textile and garments, metal work and engineering, canvas, food processing and beverage, leather and leather products, soft drink), agriculture (food crops, coffee, milk, hen, irrigation), construction (mining, engineering activities,), services (transport, cafeteria, tourism, maintenance, electronics and software, project work).

According to [5], before going to FeMSEDA of Ethiopia to begin the MSEs procedure, it is advisable to consider the following points: Decide on the MSEs sector of interest, calculate the MSEs capital required, have an idea of member's requirement and have an idea of Environmental situation. That is, Getting MSEs advice is a critical issue for members' since knowing the right area is a key factor and also knowing the right enterprise to implement in is another factor to consider for new members [10]. Members tend to lose money by making the

Wrong choice MSEs sector and choosing the wrong enterprises.

But in Ethiopia, most MSEs members consider only their interest without considering their capitals, location of enterprises, educational levels, age, gender, infrastructures needed in that specific location, whether there is customer in that location or not, the availability of raw materials to run the Enterprises [10].

Generally, the goal of this study is to develop a recommender system in order to give recommendation to MSEs Sector new members that will be most successful in the future based on considering the above stated characteristics.

## **1.2. Statement of the Problem**

As part of industrial sector, MSEs are increasingly becoming popular and important in the Ethiopian economy. These sectors would play a decisive role in contributing to employment generation, poverty reduction and opening distribution of wealth and opportunities. However, MSEs have faced a number of problems: among these, lack of access to market, finance, business information, lack of business premises, low ability to acquire skills and managerial expertise, low access to appropriate technology and poor access to quality business infrastructure [3], [12].

In the MSE sectors experts faced problems like, lack of appropriate, relevant and understandable information system that they need to provide advice and guidance to their members. Advice is one of the most important issues of small enterprises, in particular MSEs [14]. Lack of access to appropriate information is encountered by MSEs due to the fact that information system is not developed to enable proper collection, organization and dissemination in the sector. Even if, there are federal and regional micro and small enterprises agencies in Ethiopia, there is no integration or collaboration between different experts that are found in different MSEs to develop an organized guidance to new members. Because, collection of ideas from different members is important to develop well defined and organized guidelines to new members. For instance, one expert may have awareness about wood work related sectors but have no idea about agriculture, engineering, technology, services etc. These show that expert's advice on only the areas they are familiar with.

Difficulty of getting MSEs advice is a critical issue for members. Since knowing the right MSE sector is a key factor and also knowing the right sector to start in is another factor to consider for new members [13]. Members tend to lose money by making the wrong choices or choosing the wrong enterprises without considerations of their demographic characteristics, location of MSEs sector and their income level.

According to Ethiopian MSEs sector experts, members have a few factors to consider when looking for the right place to put their money such as safety of capital, current income and capital appreciation. MSEs Decision will also depend on a Member's age, capital, gender, educational level, location (region, zone and woreda) affecting the success of members in the activities of MSEs. The Ethiopian MSEs sector experts further comments that, members are confused about where to start the enterprise sector, what fund to use and which MSE Sector is best to them to be successful due to lack of understanding the businesses.

Therefore, as shown above one of the major problems faced in Ethiopian MSEs is the difficulty of getting speedy, consistent, clear and organized expert advice on each enterprise sectors and activities. The other problem is shortage of skilled experts and consulting service of enterprise sectors who can give advice on MSEs issues in the country, Ethiopia. This study therefore, aims to design a case based recommender system that helps members to get rapid and reliable advisory service to select new enterprises sectors. Based on this advice members can identify the areas of MSEs that have highest potential of achievement and that match their personal characteristics and potentials.

Accordingly, this study attempts to explore and answer the following research questions:

- What are the major attributes that have more influence in MSE sectors and activities selection in Ethiopia?
- What is the availability of domain experts in MSEs to advice members on different enterprise sectors and activities selection?
- How does one acquire, model, represent, and implement the required knowledge for the system that is proposed?
- To what extent does a case based recommender system provide acceptable performance in the MSE sector and activity selection in Ethiopia for new members?

### **1.3. Objective of the Study**

The study has the following general and specific objectives:

#### **1.3.1. General Objective**

The general objective of the study is to explore case based recommender system that can give recommendation to new members in the selection of MSE sectors and activities in Ethiopia.

#### **1.3.2. Specific Objectives**

With the above general objective this research tried to fulfill the following specific objectives.

- To review different global and local researches that can help for the successful completion of the study.
- To understand the concept of case based recommender system and the designing process.
- To identify and collect the previous cases, facts and rules that new members need to know in order to select MSE sectors.
- To identify main attributes that influence MSE Sectors selection of new members.
- To collect different expert ideas and develop guidelines or advice systems to new members.
- To develop a prototype recommendation system to enterprise members on the selection of MSE Sector that best matches with their characteristics or attribute.
- To evaluate the performance of the prototype system
- To forward recommendations for further work.

## **1.4. Methodology**

To realize the main goal of this study, different methods and techniques were used.

### **1.4.1. Literature Review**

For proper understanding of the problem related to MSEs, to collect necessary information related to recommender system and for successful completion of this study relevant literatures such as books, journals, magazines, conference papers, manuals, and resources from internet were reviewed.

### **1.4.2. Method of Data Collection**

In this study to acquire the needed knowledge, both primary and secondary data collection methods were employed. The main data sources for this study were the previously recorded MSEs members' cases from the FeMSEDA office in Ethiopia. The previous members cases include: amount of capital to start, number of members, locations (region, zone, woreda) that the MSEs are implemented, gender, age, status of previous members (start up, growth and maturity), the name of enterprise sectors and activities.

### **1.4.3. Sampling Technique and Sample Size**

Domain experts have been selected by using purposive sampling techniques. I.e. Based on their educational qualifications related to the domain area and their immediate position in the enterprise sector promotion 6 (six) domain experts. Because of the difficulty of interviewing all members all over the country was difficult, ten (10) enterprise members were interviewed from each region. Because, ten (10) members are represented for each regions (one member from one region) to report the progress and problems regarding to enterprises development. This interview of domain experts and members are used for asking about advising system and asking about all information gathered from secondary sources. Semi structured interview was used for both members and experts. A semi-structured interview is an interview which has a guide that usually includes both closed-ended and open-ended questions. It is more flexible than structured one. In

these kinds of interview the interviewer has a chance to change the order of questions and expand the dimension of questions based on the participants' responses.

The researcher used purposive data selecting methods to select sample datasets got from FeMSEDA of Ethiopia for this research. Because there were large amount of data and the data includes all members in different status namely in startup (infant stage which needs to start an enterprise), growth (have a power to compute on the market) and maturity stage (it is profitable and have try to add other sectors by themselves).

Therefore, to select successful enterprise sectors from each activity, the researcher selects enterprise sectors that are at growth and maturity stages as a case base development. These successful enterprise activities were selected from each enterprise sectors. There were totally 50,000 growth and matured enterprise sectors from the total of 100,000 start up, growth and maturity levels of enterprise sectors. But since the data set having different problems such as missing value of majority attributes, redundancy, etc the researcher selects five thousand one hundred thirty three (5133) growth and matured cases from the total of 50,000 cases. Finally these 5133 cases represented as case base in a CSV (Comma Separated Values) format that are used as previously solved cases. From 5133 cases used for this research 2,856 were from manufacturing, 1, 143 were from trade sector, 785 from service sector, 266 from construction sector, 83 agriculture sectors, the number of samples from each enterprise sector depends on the total number of cases in the case base. So the researcher selects the case from each sector proportionally. (*This is explained in chapter 3 table 3.1 in detail*)

#### **1.4.4. Case Representation**

After the required knowledge is acquired, the next task is knowledge (case) representation. Basically, CBR (case-based reasoning) is a problem solving methodology that addresses a new problem by first retrieving a past, already solved similar case, and then reusing that case for solving the current problem [11].

Case representation in case-based reasoning makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes.

The case representation task is concerned with the selection of relevant attributes, the definition of indexes and structuring the knowledge in a specific case implementation [7].

For this research, the acquired cases were represented using one of the different case representation methods that are appropriate for the research. Among the different case representation methods like relational database case representation, predicate based representation and soft computing case representation methods, feature-value case representation method is used. But, for this study feature-value representation method is appropriate. Feature-value case representation method is a process of representing a case as a vector of attribute-value pairs. The reason for representing cases using feature-value representation is that this approach supports nearest neighbor retrieval algorithm and it represents cases in an easy way [15].

#### **1.4.5. Implementation Tool**

Nowadays, there are several important CBR implementation frameworks available for free for teaching and academic research. Two of them are myCBR which is developed by German Research Center for Artificial Intelligence and jCOLIBRI which is developed by Group of Artificial Intelligence Applications (GAIA). Both tools have their own capabilities and limitations. For example, the main limitations of myCBR are: It does not support full CBR cycles (only Retrieval and Retain are supported) and Applicable for simple CBR applications.

So for this research jCOLIBRI CBR Framework is selected based on the following unique characteristics’.

- jCOLIBRI supports the full CBR cycles (Retrieval, Reuse, Revise and Retain discussed on chapter two)
- Suitable for developing large scale applications.
- jCOLIBRI is extensible, reusable
- It is more user friendly than others and
- It works well in external database such as plain texts. [7], [8], [16].

#### **1.4.6. Testing Procedures**

Domain experts and members were selected to evaluate the developed system. Test cases were selected by the researcher and relevant cases were selected by experts and members to test

selected test case by entering their own query. Recall and precision value of the system have been calculated based on its retrieval results. In addition to that, user acceptance testing was performed to test its application by the user and relevant feedbacks were collected to calculate its performance.

### **1.5. Scope and Limitations of the Study**

Due to different factors such as limitation of time and resources, the scope of the study is restricted only on developing prototype recommender system for enterprise sectors and activity selection based on different characteristics of the members. The recommender system does not include other activities of the MSEs agency such as constraints of finance, infrastructure problems, giving and renewal of enterprises license, calculation of taxation, etc.

The main limitation in this developed system is: it does not control the dynamic nature of the environment. As environmental conditions are changed day to day, with social context, technological factors and socio economical contexts this system does not have a capacity to control them. These factors may affect the establishment as well as success of enterprise sectors'.

The other limitation of the study is: the case structure for the prototype is made up using attributes that are recorded in the file of previous members. Due to these attributes like: land size requirement or enumeration area to start enterprises, marital status, and level of risk taker which are considered in enterprise sector and activity selection are not included in the case structure for the prototype. Other limitation of the prototype is, the explanation facility does not give response or feedback based on member's questions. It gives only a direct explanation when the system recommends a solution. For instance, if one member not clear about enterprise type, there is no a possibility to ask explanation.

### **1.6. Significance of the Study**

Mainly Ethiopian MSEs development agency experts and new members, specifically the inexperienced MSEs expert professionals are the immediate beneficiaries to enhance their day to day activities. The prototype has great significance to teach new MSEs experts and members in

order to have well understanding about MSEs success and failure. As a result, those experts of MSEs can use the system in recommending successful MSE sectors to new members based on their personal or socio economical characteristics. The system can give better advice services where there are no highly qualified MSEs expert professionals. Because, the case based recommender system is developed using the knowledge of different sector of MSEs domain experts and written sources.

## **1.7. Organization of the Thesis**

The study is organized into six chapters.

**Chapter one** is the introduction part, which contains the background of the study, problem statement, objectives, scope of the study, the significance of the study and methodology to carry out the research.

**Chapter Two** discusses about Review of literature on the knowledge based systems, about its background, architecture, development phases, and knowledge based system overview and opportunities for enterprise sectors are discussed.

**Chapter three** discusses the knowledge acquisition, representation and conceptual modeling procedures.

**Chapter four** deals with implementation and design of Case based recommender system.

**Chapter five** presents the results found in the evaluation and testing process of the prototype case based recommender system.

**Chapter six** focuses on the conclusion and recommendation based on the finding of the research and recommendations are proposed for future research.

## CHAPTER TWO

### 2. LITERATURE REVIEW

#### 2.1. Recommender System

Recommendation systems have arisen to provide convenient suggestions to users. These systems can be used for different purposes in several domains from offering papers to researchers to helping consumers in e-commerce. There are recommendation systems in different domains such as films, television programs, video, music, books, news, images, and web pages. It can be said that, recommendation systems basically aim to overcome the difficulty of finding proper information. Among the most famous ones, Amazon recommends books in book domain; Last.fm helps users to find the songs that they want to listen; and Movie Lens tries to guide users to reach the movies they might like [7]. In a recommendation system application there are two classes of entities, which we shall refer to as users and items. Users have preferences for certain items, and these preferences must be teased out of the data. The data itself is represented as a utility matrix, giving for each user-item pair, a value that represents what is known about the degree of preference of that user for that item.

In the simplest recommendation process, the user is supposed to be looking for some product to purchase and therefore is asked by the system to provide some product requirements, those that he/she considers as the most important. In reply, the system initiates a search in the case base to identify products that should be recommended, i.e. those that satisfy these requirements. In this process we can identify some basic elements, such as, the input (where the user provides his/her requirements), the products retrieval (where the system searches the products according to user requirements) and the output, where some recommendation is given to the user [11].

Recommender systems are being used in e-commerce web sites to help customers in selecting products more suitable to their needs. The growth of Internet and the business to consumer e-Commerce has brought the need for such a new technology. One of these approaches, called knowledge-based recommender system, allows the system to learn about user preferences over time and automatically suggest products that the learned user model [11].

Recommender systems represent user preferences for the purpose of suggesting items to purchase or examine. They have become fundamental applications in electronic commerce and information access, providing suggestions that effectively prune large information spaces so that users are directed toward those items that best meet their needs and preferences [19].

Recommender systems try to help user's access complex information spaces. A good example is when they are used to help users to access online product catalogs, where recommender systems have proven to be especially useful for making product suggestions in response to evolving user needs and preferences [20].

## 2.2. Architecture of Recommender system

A recommendation seeker may ask for a recommendation, or a recommender may produce recommendations to the user with no prompting. Seekers may volunteer their own preferences, or recommenders may ask about them. Based on a set of known preferences his/her own, the seeker's, and those of other people, often people who received recommendations in the past, the recommender recommends items the seeker probably will like. The seeker may use the recommendation to select items from the universe or to communicate with others [7]. The architecture of recommender system is represented under here.

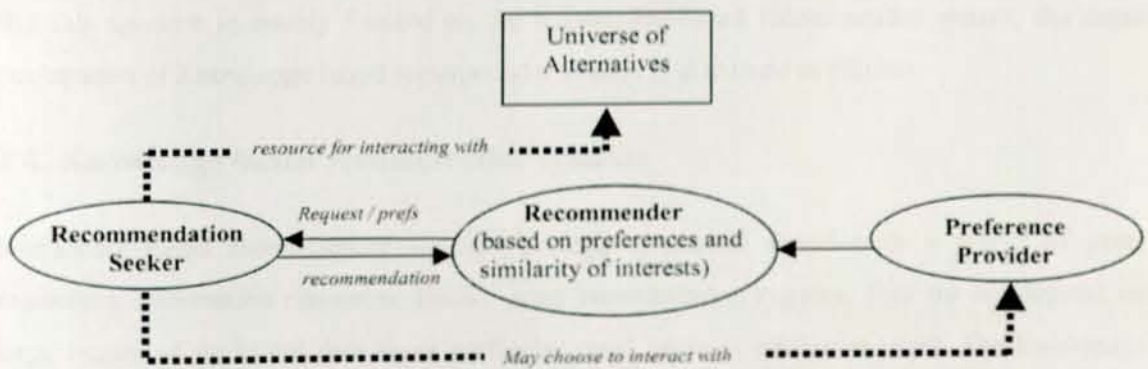


Figure 2.1: Architecture of recommender system (source: [8])

As shown in figure 2.1, the recommendation seeker (members) asks recommendation with the help of recommender system from preference provider and universe of alternatives by entering the required queries by them. Based on the items listed in the preference provider, the system

provides solution or recommendation by measuring similarity between new cases which is provided by the members and existing case stored in the case base. Finally, the system will provide best similar cases to the members from the existing cases.

### 2.3. Types of Recommender System

Based on how recommendations are made, Recommender systems can be classified into the following categories [9]:

- ✓ **Content-based recommender systems:** in which the user is recommended items similar to those the user preferred in the past.
- ✓ **Knowledge based recommender system:** in which it depends either on explicit domain knowledge about the items or knowledge about the users to derive relevant recommendation.
- ✓ **Collaborative filtering systems:** in which the user is recommended items that people with similar tastes and preferences liked in the past.
- ✓ **Hybrid recommender system:** Due to the shortcomings proper of each of these strategies alone, combinations of content based and collaborative filtering have been investigated in the so called **hybrid recommender systems**.

But this research is mainly focused on the Knowledge based recommender system, the detail explanation of Knowledge based recommender system is discussed as follows:

### 2.4. Knowledge based recommender systems

Knowledge-based recommender systems perform a needed function in a world of ever-expanding information resources. Unlike other recommender systems, they do not depend on large bodies of statistical data about particular rated items or particular users. The knowledge component of these systems need not be prohibitively large, since we need only enough knowledge to judge items as similar to each other [6].

Further, knowledge-based recommender systems actually help users explore and thereby understand an information space. Users are an integral part of the knowledge discovery process, elaborating their information needs in the course of interacting with the system. One need only

have general knowledge about the set of items and only an informal knowledge of one's needs; the system knows about the tradeoffs, category boundaries, and useful search strategies in the domain [6].

## **2.5. Knowledge based system (case based reasoning system)**

Knowledge-Based System (KBS) is one of the major family members of the AI group. With availability of advanced computing facilities and other resources, attention is now turning to more and more demanding tasks, which might require intelligence. The society and industry are becoming knowledge oriented and rely on different experts' decision-making ability. KBS can act as an expert on demand without wasting time, anytime and anywhere. KBS can save money by leveraging expert, allowing users to function at higher level and promoting consistency. One may consider the KBS as productive tool, having knowledge of more than one expert for long period of time. In fact, a KBS is a computer based system, which uses and generates knowledge from data, information and knowledge [21].

Case-Based Reasoning (CBR) is one of the most successful machine learning methodologies that exploit a knowledge-rich representation of the application domain. Basically, CBR is a problem solving methodology that addresses a new problem by first retrieving a past, already solved similar case, and then reusing that case for solving the current problem. In the most straightforward application of CBR to recommendation generation, the case base models the products to be recommended and the set of suggested/recommended products is retrieved from the case base by searching for products similar to that partially described by the user [20].

### **2.5.1. Architecture of knowledge based system**

The KBS consists of a Knowledge Base and a search program called Inference Engine (IE). The IE is a software program, which infers the knowledge available in the knowledge base. The knowledge base can be used as a repository of knowledge in various forms. This may include an empty Work Space to store temporary results and information/knowledge pieces/chunks. As an expert's power lies in his explanation and reasoning capabilities, the expert system's credibility also depends on the Explanation and Reasoning of the decision made/suggested by the system. Also, human beings have an ability to learn new things and forget the unused knowledge from

their minds. Simulation of such learning is essential component of KBS. The life of KBS may vary according to the degree of such simulation. KBS may be either manually updated or automatically updated by machine (machine learning). Ideally, the basic frame of a KBS rarely needs to be modified. In addition to all these, there should be an appropriate User Interface, which may have the Natural Language Processing facility [22]. These components are shown in the following figure.

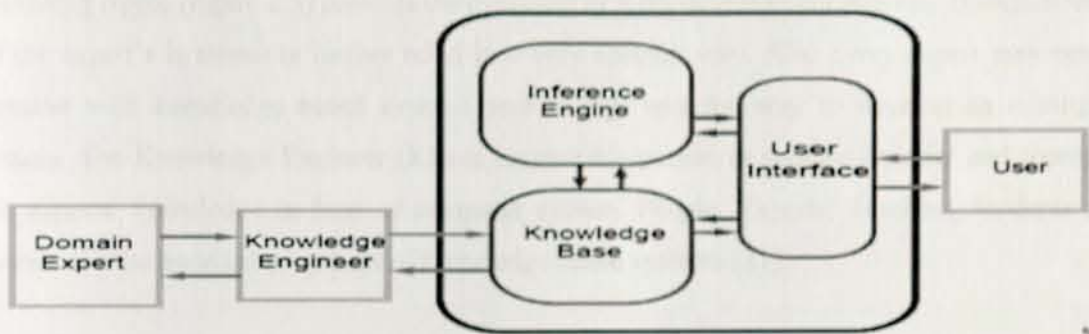


Figure 2.2: Architecture of a Knowledge-Based System (source: [22])

**Domain experts:** - individuals who currently are experts and know about the domain, but nothing about particular cases or how the system works.

**Knowledge engineers:** - design, build, and debug the knowledge base in consultation with domain experts.

**Knowledge base:** - contains necessary knowledge about the domain that is required to handle problems. Knowledge can acquire from expert, documents, books, novice and others. There are many techniques to represent knowledge such as rules, semantic network, and frames.

**Inference engine:** - also known as rule interpreter. It is a software program and problem solving component, which infers the knowledge available in the knowledge base. It allows new inferences to be made from the case specific data and the knowledge in the knowledge base.

**Explanation/reasoning:** - analyzes the structure of the reasoning performed by the system and explains it to the user.

**User:** - individuals who consults with the system to get advice with the system which would have been provided by the expert. Have problems for the system; know about particular cases, but not about how the system works or the domain.

### 2.5.2. The Knowledge Development/Engineering process

The development of knowledge based system is the integration of many components. The following figure (figure 2.3) presents the overview of KBS development process. The knowledge of the expert's is stored in his/her mind in a very abstract way. Also every expert may not be familiar with knowledge based systems terminology and the way to develop an intelligent system. The Knowledge Engineer (KE) is responsible person to acquire, transfer and represent the experts' knowledge in form of computer system. People, Experts, Teachers, Students and Testers are the main users' groups of knowledge based systems [21].



Figure 2.3: Development of a Knowledge-Based System (source: [21])

#### 2.5.2.1. Knowledge acquisition

Knowledge acquisition is the process of acquiring relevant knowledge from human experts, books, documents, sensors, or computer files. The knowledge can be specific to the problem domain or to the problem-solving procedures, it can be general knowledge (e.g., knowledge about business) or it can be Meta knowledge (knowledge about knowledge). Knowledge acquisition is the bottleneck in knowledge based system development today. Because, the

trustworthiness and the performance of the knowledge based system mainly depends upon the acquired knowledge [7].

The knowledge acquisition process incorporates typical fact finding methods like interviews, questionnaires, record reviews and observation to acquire factual and explicit knowledge. However, these methods are not much effective to extract tacit knowledge which is stored in subconscious mind of experts and reflected in the mental models [21]. The performance of the expert systems depends upon the reliability, validity and accuracy of the elicited knowledge. The process of knowledge elicitation is affected by different contributing factors such as communication between the expert and ability of knowledge engineer.

#### **i. Interview**

Is the most commonly used form of knowledge acquisition is face-to-face interview analysis. This is an explicit technique that appears in several variations. It involves a direct dialog between the expert and the knowledge engineer. Information is collected with the aid of conventional instruments (e.g., tape recorders, questionnaires) and is subsequently transcribed, analyzed, and coded. Based on its structure, interview can be classified into structured, semi structure and unstructured interview [23].

**Unstructured (informal) Interviews:** Many knowledge acquisition interview sessions are conducted informally, usually as a starting point. Starting informally saves time; it helps to move quickly to the basic structure of the domain.

Unstructured interviewing provides complete or well-organized descriptions of cognitive processes. There are several reasons for this: The domains are generally complex; the experts usually find it very difficult to express some of the most important elements of their knowledge; domain experts may interpret the lack of structure as implying that they need not prepare for the interview; data acquired from an unstructured interview are often unrelated, exist at varying levels of complexity, and are difficult for the knowledge engineer to review, interpret, and integrate; and few knowledge engineers have the training and experience to efficiently conduct an unstructured interview [23], [7].

**Structured Interviews** A structured interview is a systematic, goal-oriented process. It forces organized communication between the knowledge engineer and the expert. The structure reduces the interpretation problems inherent in unstructured interviews and allows the knowledge engineer to prevent the distortion caused by the subjectivity of the domain expert. Structuring an interview requires attention to a number of procedural issues.

A **semi-structured interview** is an interview which has a guide that usually includes both closed-ended and open-ended questions. It is more flexible than structured one. In these kinds of interview the interviewer has a chance to change the order of questions and expand the dimension of questions based on the participants' responses.

In summary, interviews are an important technique, but they must be planned carefully, and the interview results must be subjected to thorough verification and validation methodologies.

#### **ii. Observations**

Sometimes it is possible to observe an expert at work. In many ways, this is the most obvious and straightforward approach to knowledge acquisition. However, the difficulties involved should not be underestimated. For example, most experts advise several people and may work in several domains simultaneously. In this case, the knowledge engineer's observations will cover all the other activities as well. Therefore, large quantities of knowledge are being collected, of which only a little is useful. In particular, if recordings or videotapes are made, the cost of transcribing large amounts of knowledge should be carefully considered [29].

#### **iii. Document analysis**

In order to elicit knowledge from documents, relevant documents which are related to MSEs are reviewed. Document analysis collects information from existing documents. These documents include related literatures, brochures, manuals, employee hand books, reports, glossaries and training materials [8]. The documented sources used in this study are: Articles that are published in different journals, vouchers, manuals, guideline and forms that are used in the process of MSEs.

Knowledge Elicitation (KE) is the process of obtaining knowledge from a domain expert that describes how they perform a specific task and describes what general knowledge they have

about the domain. One use of KE is obtaining knowledge from a person in order to transfer it to a computer program [24].

Different KE techniques have been developed in order to obtain knowledge. They can be classified along many dimensions. The most common one used is direct versus indirect, where direct techniques, such as interviewing, are used to obtain information that is easily verbalized, and indirect techniques, such as sorting domain entities into categories, are used to obtain information that is not easily articulated in response to direct questioning. Classification can also be based on the type of interaction with the subject and the type of knowledge most commonly obtained (sequencing, classification, etc.) [24].

#### **2.5.2.2. Knowledge representation**

Acquired knowledge is organized so that it will be ready for use, in an activity called knowledge representation. This activity involves preparation of a knowledge map and encoding of the knowledge in the knowledge base [23].

#### **2.5.2.3. Knowledge validation**

Knowledge validation (or verification) involves validating and verifying the knowledge (e.g., by using test cases) until its quality is acceptable by users. Testing results are usually shown to a domain expert to verify the accuracy of the knowledge base [23].

### **2.6. Knowledge based reasoning techniques**

There are a number of knowledge based reasoning methods. The well-known reasoning approaches are ontology based reasoning, semantic network, neural network, fuzzy logic, case based reasoning and rule based reasoning. But for the purpose of this research work case based and rules based reasoning approach are discussed below in detail.

#### **2.6.1. Case based Reasoning**

Case based Reasoning (CBR) is a set of new theory and research method developed by the domain of AI in 1977 by Schank and Abelson (Schank & Abelson, 1977). The field of CBR, which has a relatively young history, arose out of the research in cognitive science. CBR is an AI

technique to support the capability of reasoning and learning in advanced decision support systems. Specifically, it is a reasoning paradigm that exploits the specific knowledge collected on previously encountered and solved situations, which are known as cases [15]. According to [16], CBR means reasoning from experiences or old cases in an effort to solve problems, critique solutions, and explaining inconsistent situations.

Case-based Reasoning (CBR) combines the knowledge based support philosophy with a simulation of human reasoning when past experience is used, i.e. mentally searching for similar situations happened in the past and reusing the experience gained in those situations. In CBR, the knowledge cases are structured and stored in a database, which the user queries when trying to solve a problem. In processing a query, the system evaluates the similarity of features between each case in the database and the query. The most similar cases are presented to the user as possible scenarios for the problem, i.e. the system doesn't make the decision it only supports the decision making process [16].

The structure of a CBR system is usually devised in a manner that reflects separate stages. However, at the highest level of abstraction, a CBR system can be viewed as a black box that incorporates the reasoning mechanism and the following external facets [26]:

- The input specification or problem situation
- The output that defines a suggested solution to the problem
- The memory of past cases, the case base, that are referenced by the reasoning mechanism

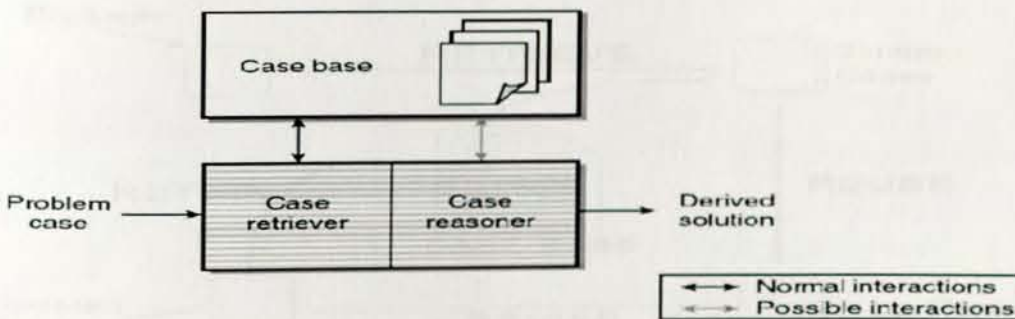


Figure 2.4: Two Major Components of the CBR System

In most CBR systems, the CBR mechanism alternatively referred to as the problem solver or reasoner has an internal structure divided into two major parts: the case retriever and the case reasoner. The case retriever's task is to find the appropriate cases in the case base, while the case reasoner's uses the cases retrieved to find a solution to the problem description given. This reasoning process generally involves both determining the differences between the cases retrieved and the current case, and modifying the solution to reflect these differences appropriately. The reasoning process may or may not involve retrieving additional cases or portions of cases from the case base [26].

### 2.6.1.1. Case based reasoning cycle

Case based reasoning life cycle incorporate four major components that make the reasoning mechanism successful. These are retrieval, reuse, revise and retain.

Solving a problem by CBR involves obtaining a problem description, measuring the *similarity* of the current problem to previous problems stored in a *case base* (or *memory*) with their known solutions, *retrieving* one or more similar cases, and attempting to *reuse* the solution of one of the retrieved cases, possibly after *adapting* it to account for differences in problem descriptions. The solution proposed by the system is then *evaluated* (e.g., by being applied to the initial problem or Assessed by a domain expert). Following *revision* of the proposed solution if required in light of its evaluation, the problem description and its solution can then be *retained* as a new case, and the system has learned to solve a new problem [25].

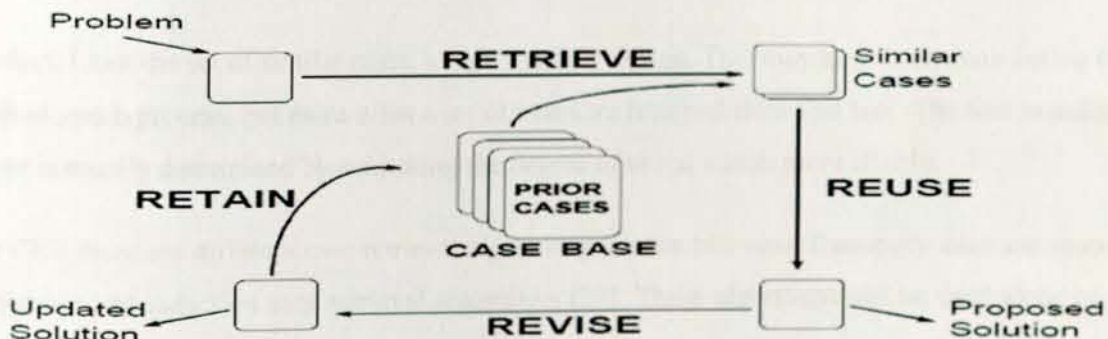


Figure 2.5: The case based reasoning cycle (source: [25])

#### 2.6.1.1.1. Retrieval

An important step in the CBR cycle (Figure 2.5) is the retrieval of previous cases that can be used to solve the target problem.

Case retrieval is the process of finding within the case base those cases that are the closest to the current case. To carry out case retrieval there must be criteria that determine how a case is judged to be appropriate for retrieval and a mechanism to control how the case base is searched. The selection criteria is necessary to decide which case is the best one to retrieve, that is, to determine how close the current and stored cases are [26].

The Retrieve task starts with a partial problem description, and ends when a best matching previous case has been found. Its subtasks are referred to as Identify Features, Initially Match, Search, and Select, executed in that order [27].

**Identify Features:** To identify a problem may involve simply noticing its input descriptors, but particularly for knowledge-intensive methods a more elaborate approach is taken, in which an attempt is made to understand the problem within its context. Unknown descriptors may be disregarded or requested to be explained by the user. In other words, it identifies its descriptive properties and takes out the properties which don't describe the problem strongly.

**Initially match:** An initial matching process which retrieves a set of plausible candidates, and a more elaborate process of selecting the best one among these. That means it involves searching and similarity assessment to produce similar cases.

**Select:** From the set of similar cases, a best match is chosen. This may have been done during the initial match process, but more often a set of cases are returned from that task. The best matching case is usually determined by evaluating the degree of initial match more closely.

In CBR there are different case retrieval algorithms but the two most frequently used are nearest neighbor and induction case retrieval algorithms [27]. These algorithms can be used alone or in combination with each other.

#### 2.6.1.1.2. Nearest Neighbor Retrieval

In nearest neighbor retrieval, the case retrieved is chosen when the weighted sum of its features that match that query case is greater than the other cases in the case base. In simple terms, a case that matches the query case on  $n$  number of features, will be retrieved rather than a case which matches on  $k$  number of features where  $k < n$ . Some features that are considered more important in a problem solving situation may have their importance denoted by weighting these features more heavily in the matching [26].

In short, when a new case doesn't exactly match with old cases then this algorithm will return nearest match from CBR library. It is suitable when there are attributes that have numeric (continuous) value. The algorithm for nearest neighbor is as follows [15]:

- *For each feature in the input case*
- *Find the corresponding feature in the stored case base*
- *Compare the two values to each other and compute the degree of match*
- *Multiply by a coefficient representing the importance of the feature to the match*
- *Add the results to drive an average match score*
- *This number represents the degree of match of the old case to the input.*

#### 2.6.1.1.3. Inductive approaches

When inductive approaches are used to determine the case base structure, that is to determine the relative importance of features for discriminating between similar cases, the resulting hierarchical structure of the case base provides a reduced search space for the case-retriever [26].

In inductive retrieval, use past cases to extract rules or construct decision [15]. This technique finds target case-based on index source case. Cases are divided into a decision tree structure. Inductive retrieval used to retrieve set of matched cases and then nearest-neighbor retrieval rank these cases according to their similarity with target case.

#### **2.6.1.2. Reuse**

Once a matching case is retrieved, a CBR system attempts to reuse the solution suggested by the retrieved case in order to derive the solution for the new problem situation by adapting the retrieved case if necessary to completely fulfill the requirements of the present situation [28].

According to [27] There are two main ways to reuse past cases: (i) reuse the past case solution (transformational reuse), and (ii) reuse the past method that constructed the solution (derivational reuse). In transformational reuse the past case solution is not directly a solution for the new case but there exists some knowledge in the form of transformational operators such that applied to the old solution they transform it into a solution for the new case .

Derivational reuse looks at how the problem was solved in the retrieved case. The retrieved case holds information about the method used for solving the retrieved problem including a justification of the operators used, sub goals considered, alternatives generated, failed search paths, etc.

#### **2.6.1.3. Revise**

Depending on the employed adaptation procedure, the correctness of the suggested solution often cannot be guaranteed immediately. Then it becomes necessary to revise the solved case. How such a revision is performed, strongly depends on the particular application scenario. For example, it might be possible to apply the suggested solution in the real-world to see whether it works or not. However, often a direct application of an uncertain solution is impossible due to the corresponding risks (e.g. medical diagnosis systems). Then the revision has to be performed manually by a human domain expert or by alternative methods such as computer simulation. Usually, the focus of the revise phase lays on the detection of errors or inconsistencies in the suggested solution and the initiation of further problem-solving attempts [30].

#### **2.6.1.4. Retain**

Retention is the process of incorporating what is valuable to retain from the new problem solving episode into the existing case base in order to make the knowledge available for future reuse. This step completes the experience feedback loop that is a necessary prerequisite for enabling a system to learn from experiences [30].

This is the process of incorporating what is useful to retain from the new problem solving episode into the existing knowledge. The learning from success or failure of the proposed solution is triggered by the outcome of the evaluation and possible repair. It involves selecting which information from the case to retain, in what form to retain it, how to index the case for later retrieval from similar problems, and how to integrate the new case in the memory structure [27]. These approaches often imply a reorganization of the entire case base when adding a new case, for example, by removing other cases.

Generally, the capability to acquire new case knowledge during a CBR system's lifetime principally adds these systems to the class of learning systems. Conversely, many CBR systems developed so far do not exploit this concept of the CBR cycle at all. This holds true especially for the commercially employed systems. Further, the original idea of the CBR cycle focuses on learning case knowledge.

## **2.7. CBR Techniques**

According to [28] there are different techniques of CBR, among these: Case Representation, Indexing, Storage, Retrieval and Adaptation are the commonly used techniques in any CBR research. These techniques will be discussed in brief by the following subsections.

### **2.7.1. Case Representation**

In CBR terminology cases are usually denotes a problem situation. A case can be defined as previously experienced situation which has been captured and learnt, is referred to as a past case, previous case, stored case or retained case [29].

A case can be said to be the record of a previous experience or problem. The information recorded about this past experience will depend on the domain of the reasoned and the purpose to which the case will be put [26]. Whenever we want to choose representation format of a case, the following factors should be considered [15], [26]:

- ✓ The cases may have segments within them that form natural sub-cases or components.
  - ✓ The content or features that describe a case have associated types and structures.
- Selection of language and shell may reduce number of formats for case representation.

- ✓ Indexing and search mechanism. Format of case should be selected according to search mechanism. Case format should be able to interact with mechanism easily.
- ✓ Type or structure associated with content.

Whatever format the cases are represented in, the collection of cases itself has to be structured in some way to facilitate the retrieval of the appropriate case when queried. There are approaches such as: A flat case base is a common structure; in this method indices are chosen to represent the important aspects of the case and retrieval involves comparing the current cases features to each case in the case base. Another common case base structure is a hierarchical structure that stores the cases by grouping them to reduce the number of cases that have to be searched. As with the form of case representation chosen will also depend on a number of factors [26].

- ✓ The representation used in the case base.
- ✓ The number and complexity of cases being stored.
- ✓ The number of features that are used for matching cases during searches.
- ✓ Whether some cases are similar enough to group together

Each case is constituted with three main parts: [28]

- ✓ **Problem/situation description:** describing the state of the world when the case occurred and, if appropriate, what problem was needed to be solved at that time.
- ✓ **Solution:** describing the solution to the problem specified in the problem description or the reaction to the given situation.
- ✓ **Outcome:** the resulting state of the world when the solution was applied.

Cases are used to represent domain knowledge of a case based system. A case refers to specific experience or knowledge tied to specific situation that is worth remembering for future use. So that cases in the knowledge base represent collection of specific experienced, captured and learned situations of the application domain [30].

### 2.7.2. Indexing

Case indexing refers to assigning indices to cases for future retrieval and comparisons. This choice of indices is important to being able to retrieve the right case at the right time. This is because the indices of a case will determine in which context it will be retrieved in future. These

are some suggestions for choosing indices [26]. For example, databases use index to speed up retrieval of data. Information in a case can be two types:

1. Indexed information use for retrieval.
2. Unneeded information that may provide information to user but not used in retrieval.

For instance, we can see a medical system, where patient's age, sex, height and weight can be used as index features. That information is helpful for future retrieval. The patient's name can be included as an unneeded feature which can't be used in the retrieval. Name should be helpful for doctor for remind patient. Index should have following features [15].

- ✓ Be predictive.
- ✓ Should show the purpose for which case will be used.
- ✓ It should be easy to recognize it in the future.
- ✓ It must address the future use of the case base.

Indices should be abstract enough to allow for that cases retrieval in all the circumstances in which the case will be useful, but not too abstract. When a case's indices are too abstract that case may be retrieved in too many situations or too much processing would be required to match cases [26], [7]. Indices can be selected in both ways: manually and automatically. If you choose index manually then you have to decide a purpose of case. Automated indexing has many methods including:

- In different-based indexing, index has to differentiate a case from another case.
- In the inductive learning method, features are identified which latterly uses as an index.
- Appropriate set of indices produce by similarity and explanation-based method for those cases which have same information.

### **2.7.3. Storage**

One important aspect of the efficient CBR system is case storage [26]. It represents a logical view what is stored in case. For efficient retrieval, case base should be organized in a manageable way. These methods referred as a case-memory model.

#### **2.7.4. Retrieval**

Case retrieval is a process of finding cases which are closest to current case [26]. For efficient case retrieval, there should be selection criteria which will judge a case. CBR major research area is retrieval. There are four retrieval techniques in which all of them are discussed in the CBR cycle section.

#### **2.8. Case Adaptation**

It is a technique to alter retrieved case for reproducing new solution for new problem. A good adaptation of old cases to fit the new case can reduce significantly the amount of work needed to solve it. According to [26], there are three techniques which are mostly used in CBR systems

##### **i. Structural Adaptation**

In structural adaptation, formulas and rules are directly applied to the stored solution in CBR library. When case applies to these rules and formulas, the CBR system adapts this case and match with new problems.

##### **ii. Derivational Adaptation**

It is a technique to reuse the rules and formulas to produce a new solution to a current problem. Solutions which are retrieved must be stored as additional case in the CBR library so it reproduces new solution to the new case.

##### **iii. Null Adaptation**

It is the simplest kind of adaptation. Null adaptation means the adaptation is left to the users or it is even not necessary to adapt at all. Most commercial CBR systems work with null adaptation. The case base is searched for a similar case containing the most similar user requirements and the solution from this case is taken to solve the current problem without any modification.

#### **2.9. Advantage of case based reasoning**

CBR has several advantages in the development of knowledge base system some of them are listed as follows [36]:

- Requires less maintenance effort

- Improve problem solving performance through reuse
- Makes use of existing data, e.g. in databases
- Improve over time and adapt to changes in the environment
- High user acceptance
- Handling unexpected or missing values.

### **2.10. Disadvantage of case based reasoning**

Even though case based reasoning approaches have a number of advantages. But, due to lack of sufficient cases, the construction and inference mechanism of a case-based system loss the required objective. Some of the limitations in case-based reasoning are [29]:

- Inability to express general knowledge
- Knowledge acquisition problems
- Inference efficiency problems and Provision of explanations

### **2.11. Integrating Rule-based and Case-based Reasoning**

Rules represent general knowledge of the domain, whereas cases represent specific historical knowledge. Rule-based Systems solve problems from scratch, while case-based systems use pre-stored situations to deal with similar new instances. Therefore, the combination of the two approaches turns out to be natural and useful.

The combination of RBR (Rule Based Reasoning) and CBR is also justified by the fact that it closely emulates the human way of thinking in a variety of complex application domains (e.g. legal reasoning and medicine). When dealing with a new situation, an expert usually combines general (rule-based) knowledge with specialized (case-based) knowledge gained from experience [37]. Therefore, the integrated reasoning approach makes use of both existing knowledge and the past experiences. This integrated approach eliminates the drawbacks of each method and provides a better way to handle problems, which combine both inductive and deductive approaches.

## 2.12. CBR System Performance Evaluation Methods

Evaluation can be defined as an iterative process of systematic assessment of knowledge based system. Basically, system performance evaluation procedure tends to assess not only the technical aspects of knowledge based system but also user's satisfaction. The evaluation criteria depend up on the purposes of designed knowledge based system in the domain area [29].

Knowledge based system evaluation process involves to determine the suitability and desirability of the prototype. Effective knowledge based system evaluation process incorporates both technical and non-technical aspects. The technical aspects include exploring of the code, examining the correctness of reasoning techniques, checking the efficiency and performance of the system and debugging errors in the early age of a system development. The non-technical aspect includes system compatible with users' satisfaction, the easiness of the system, the quality of the user interface and the acceptability of the system in the real world environments [29].

As retrieval task of the CBR aims to retrieve cases relevant cases from the case base, precision and recall are useful measures of retrieval performance in CBR. Recall is defined as the ratio of the number of relevant cases returned to the total number of relevant cases for the new case in case base. Whereas precision is the ratio of the number of relevant cases returned to the total number of cases for a give new case [7], [15].

$$\text{Recall} = \frac{\text{Number of relevant cases retrieved}}{\text{Total number of relevant cases}}$$

$$\text{Precision} = \frac{\text{Number of relevant cases retrieved}}{\text{Total number of cases retrieved}}$$

## 2.13. Knowledge based system development tools

There are different types of tools that can be used for developing a CBR system. Most of these tools are commercial and few of them are non-commercial. Some of them are discussed as follows:

### **i. CBR-Express**

This CBR tool primarily designed for help desk applications by Inference Corporation. It provides a comfortable user interface and fast retrieval speed. It has simple case structure and nearest neighbor retrieval algorithm of cases. The key features of CBR-Express are to handle free-form text and enable customers to describe their problem in own words. The use of trigrams means that CBR-Express is reluctant of spelling mistakes and typing errors such as letter transpositions. It stored cases in relational database. CBR-Express is network ready and cases can be shared between organization's networks

### **ii. myCBR**

myCBR is one of the most popular CBR software platforms. It is a framework with certain capabilities and limitations. The most popular version of myCBR is a plug in of open source ontology editor Protégé (Stahl & Roth-Berghofe, 2008), but there are available some web-based versions and integration into other software. myCBR is developed by the German Research Center for Artificial Intelligence. The platform has open source code written in Java and is accessible to all users. It can be easily modified by the users depending on the purpose. The purpose of myCBR is to minimize the efforts to create CBR applications. For this normal use, without modifying the source code, no programming skills are required, but expertise in a specific CBR-developed applications. Its programming code is well documented.

### **iii. jCOLIBRI**

jCOLIBRI is a technological evolution of COLIBRI and it is an object-oriented framework in Java which is designed for building CBR systems. It is a java-based and uses JavaBeans technology for case representation and automatically generation of user interface. This framework is developed by the GAIA artificial intelligence group in Complutense University in Madrid. The framework is built in two hierarchical levels- upper and lower. The lower level consists of library of classes (Software modules) for full 4Res (retrieval, reuse, retainment and revise) CBR cycle, also for definition of cases, attributes and connectors for access to outer databases. The upper level is "black box" – graphical interface, which allows non-complicated user CBR application generation based on lower level's modules.

jCOLIBRI supports full CBR cycle. At the retrieve stage the nearest N cases are retrieved. At reuse stage several methods for adaptation are available (direct proportion and also in ontology). At revise stage methods for revision of cases are realized, as well methods for new indices generation and methods for decision making (preference elicitation). At retain stage there are methods for query retaining to the case base for future use. jCOLIBRI allows retrieval from clustered and indexed case bases and submits program interfaces (connectors) to access text and XML files, as well standard and descriptive logic databases. These interfaces can be used for diagnostic systems database access. There are lots of CBR applications, developed on jCOLIBRI based: additional shells (abstract levels) for distributed CBR systems, statistical CBR systems, multi-agent supervisor systems for text file classification, and lots of CBR recommender systems.

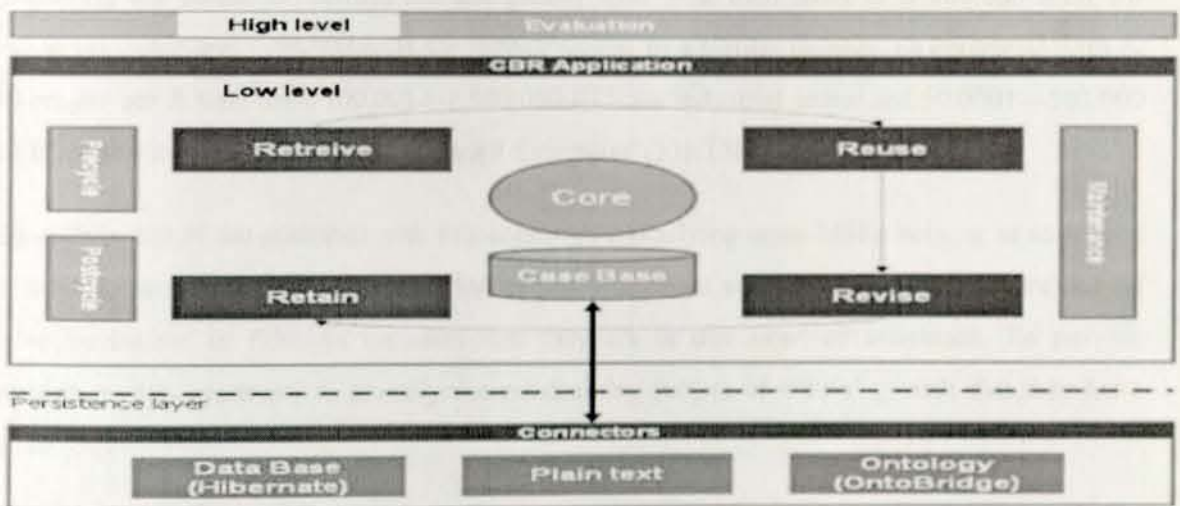


Figure 2.6: Software Architecture of jCOLIBRI

## 2.14. MSES in Ethiopia

The term MSES covers a wide range of definitions and measures, varying from country to country and between the sources reporting MSES statistics. Although there is no universally agreed definition of MSES some of the commonly used criteria are the number of employees, value of assets, value of sales and size of capital or turnover. However, the most common basis of defining MSES is number of employees. MSES play a crucial role in developing the economy

and in creating employment and they do not only provide employment and income opportunities to a large number of people, but also are at the forefront of technological innovations and export diversification [31].

In [38], an enterprise can be defined as an undertaking engaged in production and/or distribution of goods & services for commercial benefits, beyond subsistence (household) consumption at the household level. An enterprise might be owned and operated by a single household, or by several households jointly on a partnership basis or by any institutional body. For example, shoe-making, bakery, textile, apparel and garment, shoe repairing, transports, health, education and legal services are potentially profitable undertakings where enterprises exist.

An enterprise can be defined as “Micro Enterprise” when the numbers of its employees (including the owner or family) are not greater than 5 & total asset is  $\leq 100,000$  ETB for industrial sector and  $\leq 50,000$  ETB for service sector. In a similar manner, an enterprise with 6-30 employees & total asset 100,001—1,500,000 ETB for industrial sector and 50,0001—500,000 ETB for service sector is defined as “Small Enterprise”[33], [38].

Ethiopia is one of the countries with this condition and relying upon MSEs helping as an engine to drive to economic growth, elimination of unemployment and poverty and the failure rate of these businesses in Ethiopia indicates that they are in dire need of assistance. To provide assistance, it is necessary to identify the constraining factors of MSEs’ growth that they have faced [32].

The Government of Federal Democratic Republic of Ethiopia has recognized and paid due attention to the promotion and development of MSEs for they are important vehicles to address the challenges of unemployment, economic growth and equity in the country. To this effect, the government has formulated a National MSEs Development and Promotion Strategy, which enlightens a systematic approach to alleviate the problems and promote the growth of MSEs.

The Federal Micro and Small Enterprises Development Agency (FeMSEDA) Enterprise promotion efforts in Ethiopia have traditionally focused on urban based and small and micro enterprises. In the 1960s and early 1970s, a department within then Ministry of Industry and

Tourism was responsible for coordinating promotion activities which basically consisted of providing training on business management. In 1977, the Handicraft and Small Scale Industries Development Agency (HASIDA) was established to provide training mainly in management and technical skills and to serve as coordinating agency for Government policy on small enterprises. Shortage of funds and unfavorable Government policy toward the private sector in the 1980s made it extremely difficult for HASIDA to have an impact on the development of local small enterprises after that the government gives higher attention and manage the federally as well as regionally, as The Federal Micro and Small Enterprises Development Agency and the regional micro and small enterprises respectively for the following objectives.

Specific objectives of the support strategy are:

- Create long-term jobs (through skills upgrading programs for micro and small enterprises and encouraging the use of appropriate and modern technologies to improve their capacity to create employment).
- Strengthen cooperation between small enterprises.
- Promote export (especially in areas where the country has a comparative advantage).
- Provide needed technical support for the graduation of small enterprises to medium and large scale enterprises.

The Regional Micro and Small Enterprises Development Agency (ReMSEDA): The identification and selection of priority target beneficiaries will be the task of each region under the responsibility of ReMSEDA. However, broad guidelines are prepared by the federal agency to assist with the selection of the target beneficiaries. They will include micro and small enterprises which:

- Use local resources
- Have backward linkages with agriculture;
- Engage in import substitution and have the potential to export;
- Engage in activities that facilitate and promote tourism, etc.

## **2.15. Opportunity Areas for MSEs in Ethiopia**

According to [13] the Key enterprise sector opportunities for potential members exist in the following areas:

- Agriculture
- Agribusinesses and processing
- Mining and resource development
- Infrastructure development
- Manufacturing and industry and
- Construction and Tourism.

## **2.16. Related Research works**

Regarding to recommender system for the selection of enterprises sectors and enterprise activities for new members, there is no local research in our country and the researcher review foreign researches on Accelerating growth in high tech MSEs through tacit knowledge sharing using case-based recommender systems and also review local researches related to case based recommender system. Some of the related works conducted by foreign and local researchers in the knowledge based (case based) recommender system have been reviewed as follows.

Mohammad Darzi1, et al [35], has done a research on Accelerating growth in high tech MSEs through tacit knowledge sharing using case-based recommender systems the primary objective of this research was, to evaluate the fuzzy case based recommender system for knowledge sharing practice with an inter-organizational system that can facilitate the effective flow of knowledge between MSEs working in high-tech industries to speed up their growth.

According to this research work, Knowledge sharing in micro and small companies can take place in different forms like soliciting feedback, asking questions, announcing plans before execution, teamwork, speaking about what and why to do, taking advice, asking about management or employee opinions, and probing in differences of coworkers in selecting solutions. Many of them are forms of interactions between two or more MSEs and can be executed more efficiently using appropriate strategies. For instance, some of known strategies

are e-learning implementation, networking and clustering of companies, or integrative planning that comes from systems thinking methods.

The system was implemented with C# in Microsoft visual studio 2008 environment and the research attempts to perform knowledge sharing practices with employees and organization of MSES by using training and experiences and it used the proposed system by fuzzy case based recommender system finally it used to evaluate the system with the comparison of employees opinions but there is no value put in percentages.

Biazen [8], has done research study on application of case based recommender system in field of study selection in the case of higher education in Ethiopia. The objective of the author was to develop a prototype case base recommender system that assists the students in their field of study selection process. The system provides recommendation to the students based on previously solved cases and new query given by the student. The author uses 105 cases which are collected from successful students as case base. These cases are used as an input for the system to provide recommendation. This recommendation enables students to make decision easily. In this study, the author used jCOLIBRI case base development tool to develop the prototype of case based recommender system because jCOLIBRI contains both user interface which enables students to enter their query. After developing the prototype of the system, testing of the prototype for case base recommender system was done to evaluate the performance of the system. Based on user acceptance of prototype testing, the average performance of the system is 77.2% and 80.2% by the domain experts and students respectively.

Getachew [15], has done research study on application of case-based reasoning for anxiety disorder diagnosis. The main goal of this research is developing a prototype case-based reasoning system that can give decision support for anxiety disorder diagnosticians at a different level of expertise. Overcoming the limitations of a rule-based knowledge base system such as incremental learning and specific knowledge acquisition are the instigation of this research. For the implementation of the prototype, successfully solved cases are acquired from Amanuel Mental Specialized Hospital. In addition, the main parameters are identified in consultation with anxiety disorder experts. Then, the implementation of the prototype using jCOLIBRI case-based reasoning framework is realized. Finally, testing of the prototype case-based reasoning system is

done to evaluate the performance of the system. The testing of the prototype is performed from two sides. The first one is testing in terms of precision and recall and registered 71% and 82% respectively. In addition to this, the average solution similarity using methods Leave One Out evaluator and Hold Out evaluation achieved performance of 73% and 75.5% respectively. The second one is the performance of the system is evaluated by the potential users" of the system and achieved 83.2% performance.

Yibeltal [7], has done a research study on application of case-based reasoning investment activity selection. The main objective of this research is developing a prototype case-based reasoning system that can give advice for new investors' for the selection of investment activities. For the implementation of the prototype, successfully solved cases are acquired from Ethiopian investment agency. Then, the implementation of the prototype is done by using jCOLIBRI case-based reasoning framework. Finally, testing of the prototype case-based reasoning system is done to evaluate the performance of the system. Testing in terms of precision and recall and registered 64% and 85% respectively and the accuracy of previous cases also evaluated and it measures about 87%. The following table depicts the previous researches conducted by a case based reasoning with their used methods/techniques, algorithm, programming tools, no cases used and evaluation results.

Author	Research title	Method/techniques	Algorithm	Programming tool	No. of cases used	Retrieval		Accuracy
						Recall	Precision	
Mohammad Darzi (1998)	knowledge sharing using case-based recommender in MSEs	Fuzzy logic	Not specified	C# in Microsoft visual studio 2008	Not determined	Not evaluated	Not evaluated	NE
Getachew Wassie (2012)	Case based reasoning for anxiety disorder diagnosis.	feature-value case representation	NNR	jCOLIBRI	50	82%	71%	NE
Henok Bekele (2011)	Hypertension management	Voting method	NNR	Python	45	86.1	60	88.89%

Biazen (2013)	Field of study selection	feature-value case representation	NNR	jCOLIBRI	105	85%	55%	NE
yibeltal (2013)	Investment advice system	feature-value case representation	NNR	jCOLIBRI	1344	85	64	87%
Enterprise selection (2014)		feature-value case representation	NNR	jCOLIBRI	5133	86	64	90%

Table 2.1: Previous related works

After the identification of gaps, the researcher has got a new idea i.e. Implementing case based reasoning system on new area which is enterprise sector and activity selection. In order to done this research the system adopts methods, tool and algorithm from previous researches with a modification which is appropriate for this study. But the researchers used several cases than the other researchers and have got a better result than other with its retrieval and accuracy evaluation. Finally, the proposed knowledge based system in this study is conducted to explore the applicability of case based recommender system for enterprise sector and activity selection which is advantageous for both new members and experts. The main objective of the research is to assign new members in different enterprise sectors and enterprise activity based on their personal characteristics (age, gender, educational background and socio economic characteristics (such as initial capital, location of enterprises or environment, enterprise type and form of ownership). Generally, the proposed case based recommender systems can helps to enterprise experts and members during the enterprise advice process and to recommend the appropriate enterprise sectors and activities.

## CHAPTER THREE

### 3. KNOWLEDGE ACQUISITION AND MODELING

#### 3.1. Knowledge Acquisition

Knowledge acquisition is the most important process and vital stage in knowledge based system development. How and where knowledge is obtained determines the usefulness of the system. Knowledge acquisition process has many steps. Some of them are: selecting a problem to be solved by the program, interviewing an expert, questionnaires, observation, record reviews, codifying the knowledge in some representation language, and refining the knowledge base by testing it and extending its capability [51].

Therefore, the main objective of this chapter is to collect previous member cases and model how members are arranged in the selection of enterprises. The main features regarding MSEs members and the fundamental concepts that helps to develop the proposed prototype CBR system are identified. Generally, the process of knowledge acquisition of this research encompasses some basic activities such as gathering the needed knowledge, analyzing that knowledge, identifying important concepts for enterprise sectors and enterprise activities and finally modeling them by using hierarchical structure. The proposed CBR system for this study is named as Case-Based Reasoning System for Enterprise Sector and Enterprise Activity Selection (CBRSESAEAS).

##### 3.1.1. Knowledge Acquisition from domain experts

For eliciting relevant knowledge from the domain experts, six domain experts from each enterprise sectors were selected by using purposive sampling technique. Semi-structured interview technique is used to gather the required knowledge. Accordingly, enterprise officers from each enterprise sector have been interviewed to obtain the required knowledge on the domain area.

To identify the factors that affect in enterprise sector and activity selection, the researcher raised questions to domain experts about: the main attributes having effect on enterprise sector

selection, the effect of some attributes that are found from different secondary sources to get confirmation from domain experts.

The interviews were done based on the experience of experts who can give an explanation regarding the advice forwarded to members for the selection of enterprise. Most of the experts have similar idea for the factors of enterprise sector and activity selection (such as: age, gender, capital, No. of members etc). Generally, all the experts agreed that, members mostly choose enterprise sectors based on the opinion of others without considering the risk level of enterprise activity. However, experienced and knowledgeable experts can identify the best enterprise sector and enterprise activity through considering different demographic characteristics of members, socio-economic factors and other factors.

### **3.1.2. Knowledge acquisition from Members**

Primary sources of knowledge were also collected from members that participate in different enterprise sectors. To gather the required knowledge semi-structured interview technique is used and purposely, ten members were selected for interviewing from each enterprise sectors. Because, each regions have their own representatives from members.

There are issues raised for the interview of members such as: how to get advice from domain experts, what are the problems in advising systems of FeMSEDA and do experts have enough experience for giving a brief advice on where to establish the enterprise sectors. Based on the questions raised, members responded that there are not enough and experienced experts in FeMSEDA who can give suitable advice to members'. Also the advice system is different from expert to expert because, enterprise experts mostly recommend members on the familiar enterprise sectors.

Generally, all members agree that, there should be common guidelines for all advising system of enterprise sectors all over the country. Because, advice is a critical issue for members to know the right enterprise location, know the exact enterprise sector to start, and to protect members from losing money by choosing the wrong enterprise sectors and enterprise activity.

### **3.1.3. Knowledge Acquisition from Relevant Document**

Document analysis helps to collect information from existing documents. These documents include promotional literature, brochures, manuals, reports, and training materials. And also the documentary sources used in this study are: journals articles, manuals, guideline and forms that are used in the process of enterprise selection, and the Websites of FeMSEDA, (Www.mse.org.et). The main data sources for this research are previous members' cases used for developing CBRSESAEAS system found from FeMSEDA.

### **3.2. Attribute selection using data mining tools**

This algorithm is used by previous researchers to select their own best attributes fro the collected case base. In this research also used, in order to solve the problem on the collected data set of member's cases.

Attribute selection is used to investigate which attributes are the most important ones. In data mining task, one can get some attribute that has little or no impact on the overall dataset. In this members dataset there are many attributes in the data that are recorded. As a result, the researcher performs an attribute selection task by using Weka attribute selection algorithm.

The highest information gain value (i.e. ranking first) is the most important attribute for the selection of enterprise sectors and enterprise activities. Based on the WEKA tool the important attributes in ranking order are age, gender, educational level, region, woreda, zone, capital, form of ownership, type of enterprise, number of members, interested enterprise activity, and enterprise sector.

#### **3.2.1. Data Collection**

The main data sources are gathered from FeMSEDA of Ethiopia. Integration and preprocessing the dataset was the main task for the purpose of achieving objectives. These datasets or previous enterprise member cases which were collected are from the year 1990 E.c to 2005 E.c. Originally, the dataset consists of 16 attributes and 100,000 records, which include the relevant information concerning enterprise members, and type of enterprise sector as well as selected enterprise activity.

### **3.2.2. Data Preparation and Cleaning**

Data preprocessing helps to fill missing values; to detect outliers, to remove or correct some noisy data. In relation to this, data normalization needs to be performed. Moreover, the dataset should be prepared in the appropriate excel format. The data originally were available in PDF format. In order to use it in Weka it should be transformed to excel, then CSV or ARFF format.

Even if, the collected data set used for this research, have their own missed values, outliers etc. the researcher have got enough and appropriate data set for the implementation of prototype without making data cleaning tasks.

### **3.3. Knowledge Modeling**

Knowledge modeling is the representation of information in the form of logic for the purpose of processing knowledge to simulate intelligence [52]. After the knowledge is acquired from different sources, the next step is organizing and structuring of knowledge.

To make the acquired knowledge reasonable for knowledge representation, there are different conceptual modeling techniques. For this study hierarchical structure is used to model how enterprise sector and activity selection is performed by using jCOLIBRI programming tool.

After collecting the core concepts from domain experts, members and documents, the model was built by the researcher. The researcher discuss with domain experts to validate the knowledge that is acquired from different source is correct for enterprise sector and activity selection. Mainly the selection of enterprise sector and activity is done by taking consideration of the attributes such as: age, gender, educational level, amount of capital, form of ownership (trade association, cooperation and private), number of members, address (region, zone, and woreda), and enterprise type (micro and small). The prototype knowledge base system is developed based on the model presented in the following hierarchical structure and follows the procedures presented in the hierarchical structure. Finally, the conceptual modeling technique is used to show how enterprise sector and activity selection is performed. Figure 3.1 shows the model proposed.

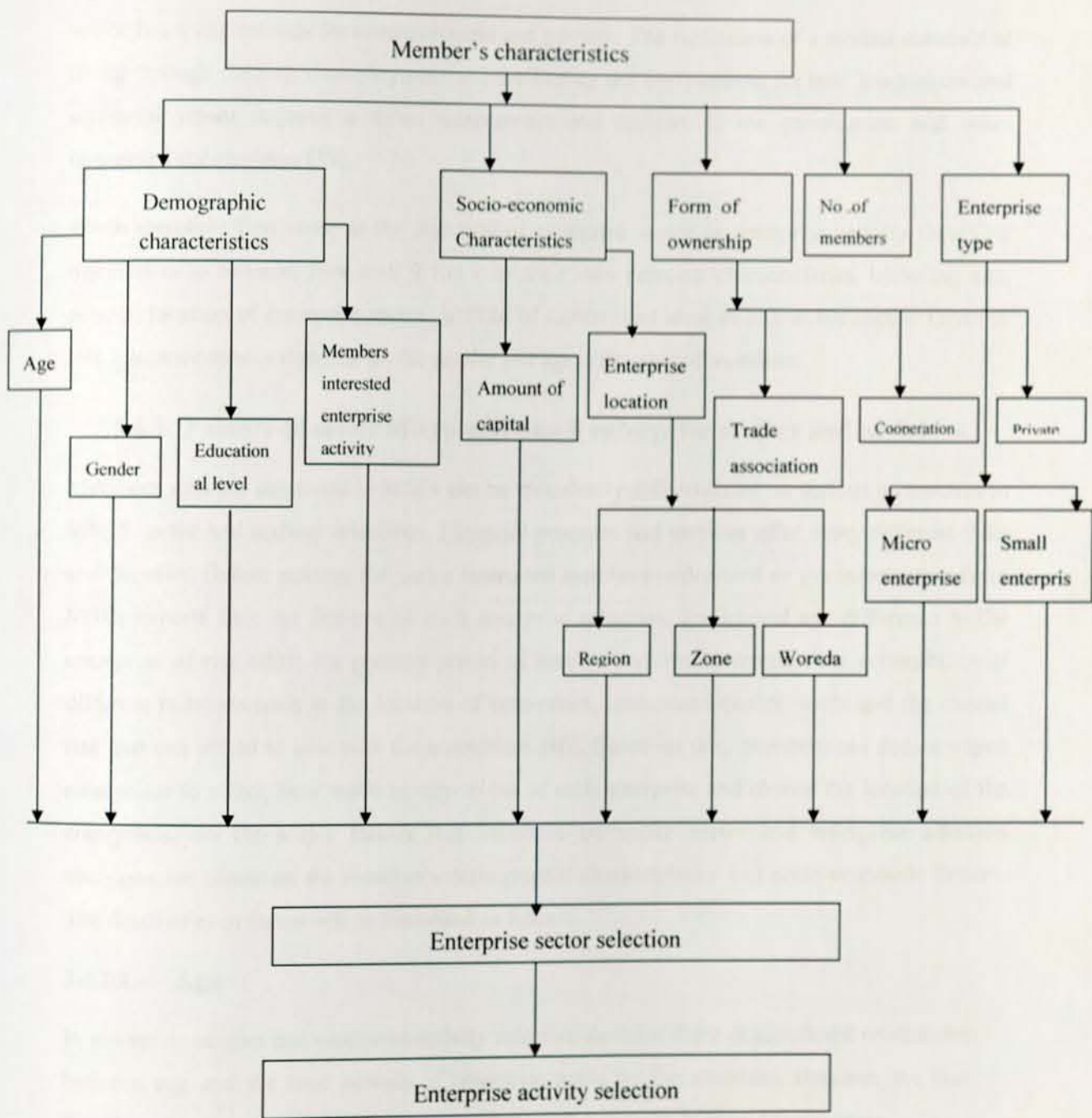


Figure 3.1: Hierarchical structure of enterprise sector selection

### 3.4. Enterprise sector and Enterprise activity selection

The MSE sector everywhere is characterized by highly diversified activities which can create employment opportunities for a substantial segment of the population. This implies that the

sector is a quick remedy for unemployment and poverty. The realization of a modest standard of living through curbing unemployment and facilitating the environment for new job seekers and self-employment requires a direct intervention and support of the government and other concerned stakeholders [39].

When members first come to the selection of enterprise sector or enterprise activity their first question is to consider how well it fits with their own personal characteristics, including age, gender, location of enterprise sector, amount of capital, and level of risk in the sector. Level of risk tolerance mostly depends on the gender and age difference of members.

### **3.4.1. Factors to select Micro and Small enterprise sectors and activities**

Members who are interested in MSEs can be specifically differentiated on various parameters in MSES sector and activity selections. Financial products and services offer many different risks and benefits. Before starting the sector interested members understand or get information from MSEs experts like: the features of each enterprise activities, gender and age difference in the enterprise of risk taker, the planned period of time to perform enterprises, the accessibility of different infrastructures in the location of enterprises, understand capital needs and the overall risk that can afford to bear with the enterprises [40]. Based on this, members can decide which enterprises to select, how much money to put in each enterprise and choose the location of the enterprises. So the major factors that influence enterprise sector and enterprise selection decisions are based on the member's demographic characteristics and socio-economic factors. The detail of each factor will be discussed as follows.

#### **3.4.1.1. Age**

In enterprise sectors and enterprise activity selection decision there is significant relationship between age and the time periods of enterprise made by the members. Because, the time duration of enterprise can vary from a few hours to few months or even several years.

Age of members in the selection of MSEs determine where to put enterprises and the life stages that affect the selected enterprise sector. The younger and middle age members (age below 50) have a chance to reach on the final financial goal and they can afford to take more enterprise risk and can be engaged on long term enterprises. Members closest to retirement

(age above 50) are likely to be more careful about taking risk because they want to protect their savings in order to have money to survive when they stop working [41]. For instance, if young members have a very long term goal, then they know that it is not very important to put in a lot of money for enterprises. On the other extreme, older aged members are thinking about retirement and just starting to save for retirement. Because of this, they are preferred to spend the maximum amount that they can afford and put their money in a relatively safe enterprise. According to [42], older enterprise members spend more conservatively than younger members and spend in short time period enterprise activity than long time period enterprise activity. There are two main classes of enterprises on the basis of period of time these are: Short-term and Long- Term enterprise. Enterprises made for a period of one to three years are termed as short-term enterprise and those that are formed for more than three years are termed as long-term enterprise [40]. Short term enterprise has its liquidity, means the ability of the firm to meet its short-term (less than a year) obligations and reveal short-term financial strength and weakness. In other words, liquidity is the ability to convert an enterprise into cash quickly. Short term enterprises are usually considered to be less risky in comparison to long term enterprise. Due to this older members spend in short term time horizon enterprise activities [42].

An enterprise sector can be formed with a partnership or in joint of members and its gender and age attribute value will be "Plc" instead of a number or male, female. Plc (private limited company) is a business arrangement in which two or more members agree to pool their resources for the purpose of accomplishing a specific task. If a partnership (plc) business loses money, the losses are divided among the members in the same way as profits. Partnership may contain a group of men and women members, young and old members with having different skills and experiences [42], [43].

The following figure shows the category of age value of members and its effect on enterprise sector selection.

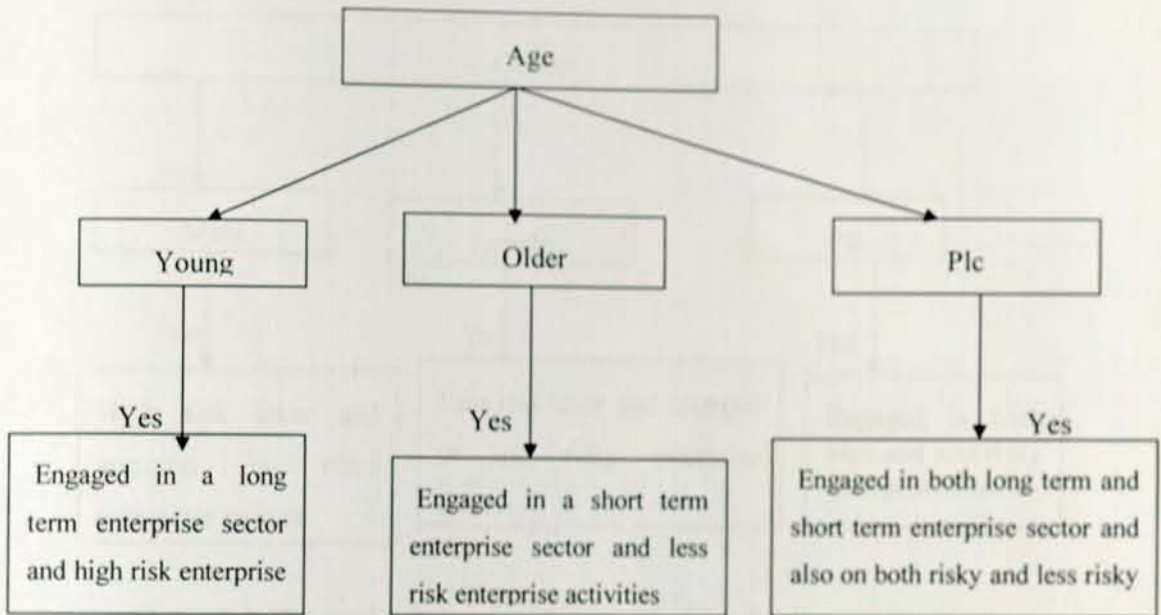


Figure 3.2: The age factors in different category of age young, old or plc

The reason for using the attribute value of age and gender as “plc” is because when members are engaged in partnership it is a group of two or more members with extremely different age and sex, so it is difficult to assign the age and gender value of all members.

### 3.4.1.2. Gender

Gender is also another factor for selecting Enterprise sectors. Women’s prefer jobs that are less risky and less competitive than men’s, and then this could explain part of the gender differences in the labor market or enterprise [43]. For example, when members are engaging in enterprise sectors, men are more often perceived to be competent, confident, put their money in capital market, and independent. While women are more often perceived to be concerned about the feeling of others, spend their funds in banking industry, risk tolerance are less likely to save in the short term as well as to save regularly, and unwilling to take a chance on losing any of their income by spend in risky assets [43]. As a result, the proportion of risk seeker is higher for male members than female members.

Members can be engaged in partnership or in joint, it is a collection of both women’s and male’s members.

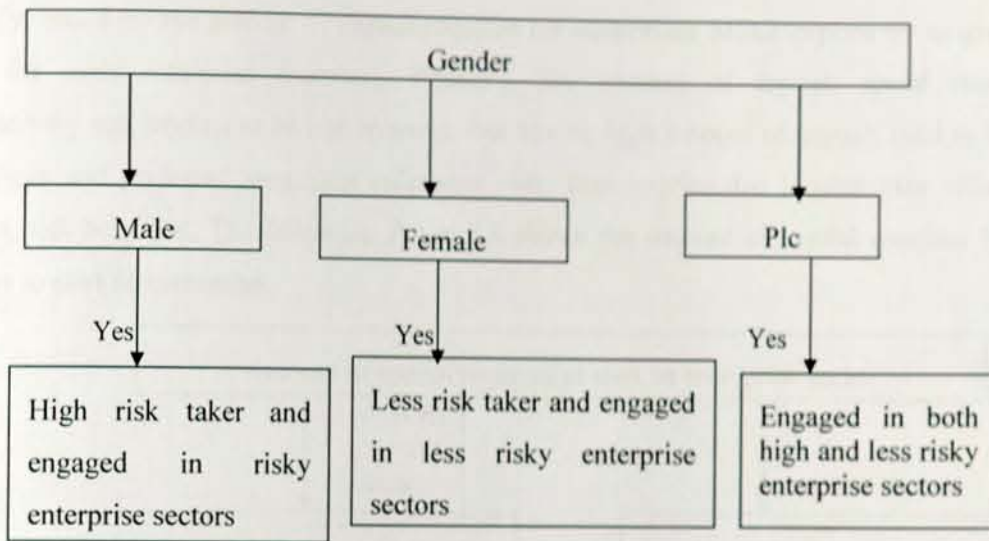


Figure 3.3: The gender factors in different category of either male or female

### 3.4.1.3. Capital

Firstly, the MSEs experts ask the new coming members for their amount of financial capitals to start the enterprises. Because, a common fatal mistake for many failed businesses enterprise is having insufficient operating capital. Members underestimate how much money is needed and they are forced to close before they even have had a fair chance to succeed. It is vital to ascertain how much money the enterprise will require (for both costs of starting and staying in business) [44].

Mostly, members with less annual capital feel less competent as a member with higher annual capital. A person with higher capital feels more successful and more powerful in daily life. According to the previous cases of MSES in Ethiopia minimum capital required to start enterprise may start from 100 birr and the maximum capital required is 500,000 up to 1.5 million birr. For example, the minimum entry capital required by members for enterprises is in areas of wholesale and retail trade, repair service sector, transport, information technology and maintenance and food related enterprises. Sectors like metal work, manufacturing and industries needs a high amount of capital to start [48].

Generally, based on the amount of capital required for enterprises MSEs experts try to give advice for each enterprise members. Because, low amount of capital: spend more conservatively and tending to be risk averters. But having high amount of capital: tend to be risk seekers and preferred long term enterprise [44]. This implies that income may affect members risk behavior. The following figure 3.4 shows the amount of capital required by members to start an enterprise.

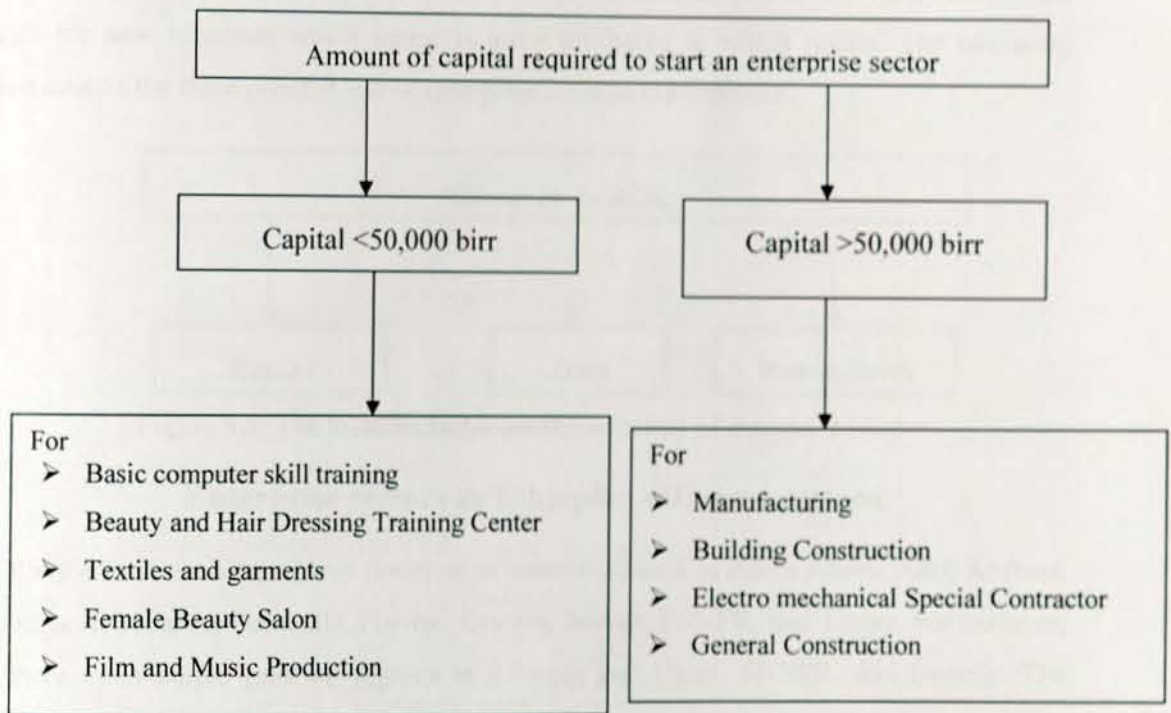


Figure 3.4: The amount of capital required for starting an enterprise sectors

#### 3.4.1.4. Enterprise Location

An enterprise location is also critical for the success of an enterprise sectors. A good location may facilitate a struggling business to ultimately grow, whereas an enterprise situated in a poor location will be at a disadvantage. The following factors are considered by the experts of MSEs when giving advice for new members at a start up stage: where the targeted customers live, the traffic, accessibility, and parking, the physical distance from competitors, access of infrastructure and the condition and safety of a building [42]. For instance, Infrastructure aids economic development of a state by creating access to regional and national markets. Throughout all the literatures, access to infrastructure has been shown to

attract members and the available land has also been cited as an important location factor [41].

In short, all the above ideas on Locations of enterprise are included or obtained in regions, zones, and woredas/town because from nine regions and two city administrations of Ethiopia. Some of the regions are suitable for agriculture, textile products while others are advantageous for mining, services sectors etc. Based on these concepts, experts try to forward advice for new members which sector is more profitable in which region. The following figure depicts the three possibilities of enterprise locations in Ethiopia.

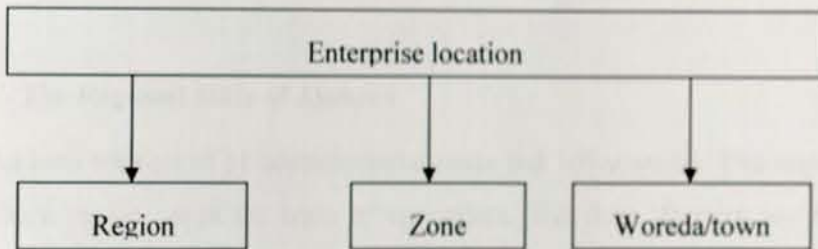


Figure 3.5: The location factor for the selection of enterprise sectors

### **Enterprise sectors in Ethiopian different regions**

In Ethiopia there are nine regions (location of enterprise) such as Addis Ababa, Afar, Amhara, B/Gumze, DireDawa, Gambella, Harire, Oromia, Somali, SNNPR, and Tigray. For instance, the three main coffee growing regions in Ethiopia are: Harar, SNNPR, and Oromia. The country has more genetic diversity among its coffee varieties other than any other countries. SNNPRS and Oromia are the most important Agricultural (honey and bees wax) producing regions in Ethiopia. Based on the previous recorded member cases, the characteristics and the major enterprise activities involve in each region are discussed as follows:

#### **I. Addis Ababa**

Addis Ababa is a city administration and has 10 sub cities. Addis Ababa is one of the major industrial zones in the country and the majority of enterprise sectors in Addis Ababa are industry (i.e. manufacturing and construction), hotel and tourism, commercial real estate, whole seal and trade, renting and business activity, educational service, health and social services, tour operation and travel agent.

## **II. Dire Dawa**

Dire Dawa is the second city administration in Ethiopia. The major enterprise sectors in Dire Dawa are in the areas of agriculture (such as lives stock, animal fattening and coffee processing), construction, textile products, business center, hotel and tourism, educational service.

## **III. The Oromia Regional State**

The State of Oromia has 12 administrative zones and 180 woredas. The major enterprise sectors in Oromia are: manufacturing (wood and metal work, textile and garments), Agriculture, services (hotels and restaurants, renting estate and business activities beauty salon), wholesale and trade and industry.

## **IV. The Regional State of Amhara**

The State of Amhara consists of 11 administrative zones and 105 woredas. The major enterprise sectors in Amhara region are in the areas of agriculture (fish farm, forestry and hen and bee production), hotel and tourism, construction and manufacturing as well as trade.

## **V. The Tigray Regional State**

The State of Tigray consists of 4 administrative zones, one special zone, and 35 woredas. The major enterprise sectors in this region are on textile and garments, wood and metal works, hotel and restaurant, mining, construction.

## **VI. The Regional State SNNPR**

The State of SNNPR is administratively divided in to 9 zones and 77 woredas. The major enterprise sectors in these regions are agriculture, hotel and tourism, wood and metal work, trade, construction and textile factory.

## **VII. The Regional State of Somalia**

The State of Somalia has 9 administrative zones and 49 woredas. The state is very rich in livestock. Moreover, it is endowed with natural gum, natural salt (in Afdem zone), natural gas oil has high potential for enterprise. The Major enterprise sectors are in agriculture, in education service and animal trade.

### **VIII. The Regional State of Harare**

Harare has no administrative zones or woredas. The total number of kebeles of the region is 19. The major enterprise sectors are construction, construction machinery rental, dairy farming, hotel and tourisms.

### **IX. The Regional State of Gambella**

The State of Gambella has two administrative zones and eight woredas. Gambella has potential enterprise opportunity in the production of cotton, sesame and other oil seeds. Fishing, mining gold and exploring petroleum, mineral water and construction materials are other important areas of enterprise in the state. So the major enterprise sectors in this region are agriculture and real estate.

### **X. The Regional State of Benishangul Gumuz**

The State of Benishangul Gumuz has 3 administrative zones and 19 woredas. Agriculture is the sector, which attracted majority of members in this region. Farming and cattle breeding, mining (gold and silver) are predominant enterprise activities in the region.

### **XI. The Regional State of Afar**

The State of Afar consists of 5 administrative zones and 29 woredas. Agriculture such as production of maize, beans, sorghum, cotton, papaya, banana, and orange are the main enterprise activities. And also business and commercial activity, especially salt production, is another area of enterprise.

The following table shows enterprise sectors taken from each region for the development of case base development for this study.

<i>No.</i>	<i>Region</i>	<i>Enterprise sector</i>	<i>Total number</i>
1	Tigray	Manufacturing	700
		Trade	600
		Service	178
2	Amhara	Manufacturing	1296
		Trade	200
		Service	166
3	Oromia	Manufacturing	423
		Trade	446
		Service	277
		Construction	76
4	Somale	Manufacturing	24
		Trade	23
		Service	13
		Construction	14
		Urban Agriculture	24
5	B/gumuz	Manufacturing	17
		Trade	11
		Service	17
		Construction	6
		Urban Agriculture	21
6	SNNP	Manufacturing	107
		Trade	18
		Service	111
		Construction	24
7	Gambella	Manufacturing	7
		Trade	6
		Service	10
		Construction	2
		Urban Agriculture	10
8	Hararie	Manufacturing	11

		Service	12
		Construction	24
		Urban Agriculture	6
9	Adiss Ababa	Manufacturing	146
		Trade	11
		Service	29
		Construction	9
		Urban Agriculture	18
10	Drie dawa	Manufacturing	15
		Construction	5
		Urban Agriculture	20
		<b>Total</b>	<b>5133</b>

Table 3.1: Regions and enterprise sectors

#### 3.4.1.5. Number of members

In this factor what is expected from the experts of the enterprise is that tell the advantage of many and small number of members for enterprises. Larger enterprises need many numbers of members but small enterprises need small number of members. In order to start an enterprise activity an accurate amount of members with the necessary skills and credibility for their given field are needed. If the enterprise is overstaffed with many unproductive members, they would be wasting money for unnecessary labor costs. On the other hand, If there are too few members at hand, then the workload would be overwhelming for the staff to handle. As a result, the enterprise overall performance may suffer. The smaller size of an enterprise is, the more frequently members work part-time but the larger sized enterprises may contain mostly permanent members [46]. Generally, member is the largest expenditure for many enterprise sectors.

#### 3.4.1.6. Educational level

The educational level of members is also other determinant factors for enterprise selection which are the human capital determinants. Educational level of the member has an impact on

the growth of his/her enterprise. Education is defined as the number of grades completed in formal education [45].

The level of education among the member's increases with the size of the enterprise, that means members of small sized MSEs are less educated than those of larger MSEs. This indicates that, a direct relationship between the share of men and women with a high level of education and the size and growth of enterprises [46].

The human capital is represented by the following three variables: first, education which is grades completed in all types of formal education. Secondly, formal technical or vocational education, that is related to the current activity of the member. Third, training and experience, that is related to the current activity.

Generally, these points are advised by experts for members such as: the performance of the enterprise is positively influenced if the member is experienced in the sector he/she is presently engaged. For instance, one of the members may engage on an enterprise at 13 years old as a trainee in a printing house and established his own workplace at the age of 30 in the same sector. The need for professional training is more expressed by the members in the manufacturing sector compared to the ones in the trade and service sectors. The member starting up his career at an early age and especially in the same sector has been very effective on his present success in addition to formal education.

Finally, the positive effect of this factor on the success of the enterprise usually takes two forms. Firstly, the member starts the enterprise having all the knowledge required for the activities in the related sector, which reduces the high risks involved during the first years of business. Secondly, the member has the chance to convey his previous relations with customers, raw material suppliers, etc.

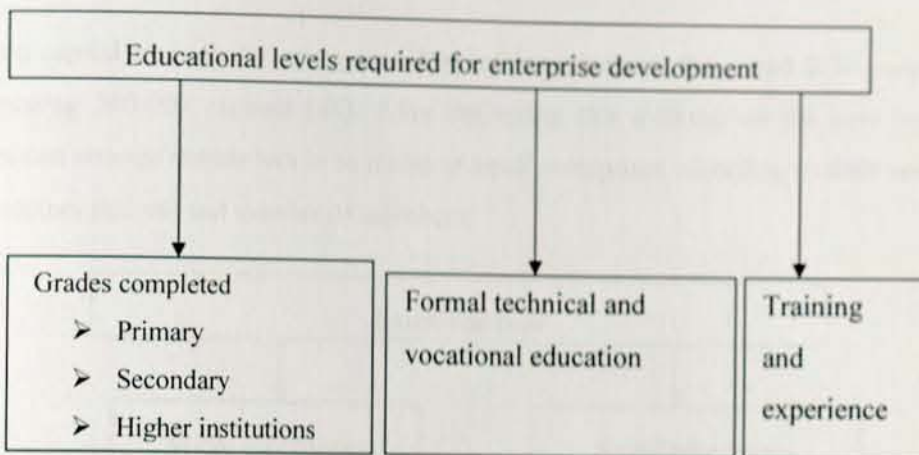


Figure 3.6: Educational level factors for the development of enterprise

### 3.4.1.7. Form of ownership

There are two forms of ownership to start up the enterprise sectors. These are individual (wholly owned) members and trade association. **Individual members** (private): means the owner of the enterprise is a single person and the responsibility of the enterprise is taken by that single person. **Trade association:** is an organization founded and funded by businesses that operate in a specific industry. **Cooperation:** means the owner of the enterprise contains more than one member and the profit as well as loss of the enterprise should be shared together [47], [48].

In this case factor, members are expected to form themselves individually or in joint form based on the features of enterprise sectors and the information given from the experts. Means there are enterprise sectors which need a high collection of members and others also need one member to start the business.

### 3.4.1.8. Enterprise type

**Micro Enterprises** are those small business enterprises with a paid-up capital of not exceeding birr 50,000. But they need up to 5 members and not exceeding 100,000 capitals for industry enterprises. For service enterprise they need 5 peoples and not exceeding 50,000 capitals [48].

**Small Enterprises** are those business enterprises with a paid up capital of above 50,000 and not exceeding birr 500,000. But they need 6-30 members and not exceeding 100,001-

1.5million capital for industry enterprises. For service enterprise they need 6-30 peoples and not exceeding 500,000 capitals [48]. After delivering this information for new members, members can arrange themselves in to micro or small enterprises according to their amount of capital, sectors interest and number of members.

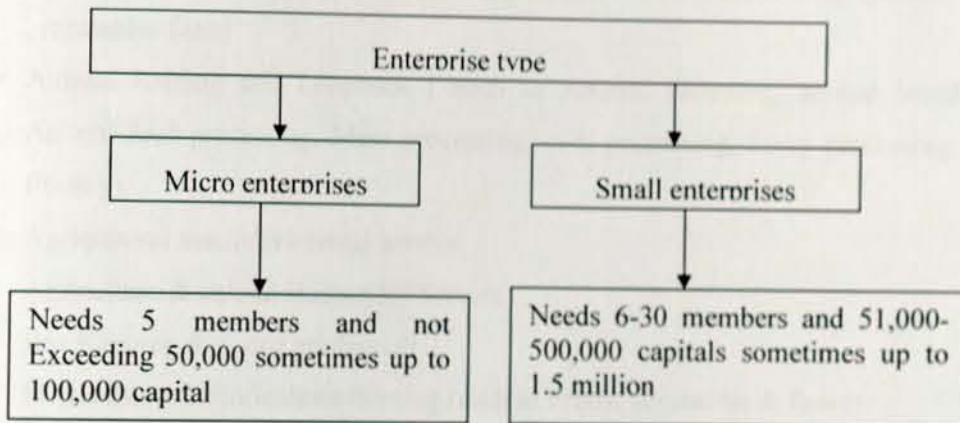


Figure 3.7: enterprise type factor for enterprise selection

### **Enterprise sectors and their respective enterprise activity**

#### **Major List of enterprise sectors are: [5]**

- Agriculture, hunting, forestry and irrigation
- Construction sector
- Educational sector
- Hotels and restaurants sector
- Manufacturing sectors
- Mining and quarrying sector
- Other community, social and personal service activities sectors
- Real estate, renting and business activities sector
- Transport, storage and communication sector
- Wholesale, retail trade & repair service sectors

## List of enterprise activities in each enterprise sectors

### 1. Agriculture, hunting, forestry and irrigation enterprise sector

- Cash crop production (such as maize, Wheat and Barley Farming, Oil seeds and pulses, Pepper ,cotton, Teff, sorghum, gum, incense, Rice Farming, Coffee, Tea ,vegetables farm)
- Animal Raising and Livestock ( such as Animal Fattening, animal breeding, Animal feed processing, Meat processing, milk processing, Dairy processing and Poultry)
- Agricultural machinery rental service
- Agriculture & animal Husbandry Service
- Bee Keeping & honey production
- Horticulture & Floriculture farming (such as Fruits, vegetables & flower.

### 2. Educational sector

The following areas are some of the potential opportunities for enterprise:

- Kindergarten, Primary, Secondary schools and high school
- colleges/universities/institutes in different field
- ICT institutions or Basic computer skill training center
- Vocational training centers
- Language School (such as Chinese, Arabic, France, Italy, English etc.)
- Art & music training center
- Athletics Training Center

### 3. Health and social work sector

The major enterprise activities in this sector are:

- Health Center & Physiotherapy Service
- Clinical laboratories and Diagnostic centers.
- Cultural Medicine Service
- Pharmacy ( animal or human pharmacy)

### 4. Hotels and restaurants sector

The major enterprise activities in this sector are:

- Hotels, Café, Bar & Restaurant
- Resorts & Lodges
- Modern Guest house, massage ,Gymnasium, steam bath service

## **5. Manufacturing enterprise sectors**

The major enterprise activities in this sector are:

- Textiles and clothing, Food and beverage products (agro- processing): Processing and preserving of meat products, fish and fish products, and fruits and vegetables etc.
- Tannery and leather goods: Tanning of hides and skins up to finished level; manufacture of luggage items, handbags, saddle and harness items,
- Paper and paper products, Building materials: Manufacture of cement, lime, gypsum, marble, granite, limestone, ceramics,

## **6. Mining and quarrying sector**

The major enterprise activities in this sector are:

- Precious & Base minerals (such as mining of gold, tantalum, platinum, nickel, potash
- Industrial & Construction minerals (such as mining and exploration of marble, granite, limestone, clay, gypsum, gemstone, iron ore, coal, copper.

## **7. Other community, social and personal service activities sectors**

The major enterprise activities in this sector are:

- Dirty water wastage
- Female Beauty Salon
- Laundry
- Sport center recreation center ( such as Children , Cinema house, bathing

## **8. Real estate, renting and business activities sector**

The major enterprise activities included in this sector are:

- Construction Material Rental
- Advertisement and printing Service
- Agricultural Machineries Rental
- Car Rent ,Car washing, loader & parking
- IT solution
- Business Center ( market center or Super market)
- Consultancy service such as in Architectural and Engineering, agricultural

## **9. Transport, storage and communication sector**

The major enterprise activities in this sector are:

- Tour operation service
- Magazine and newspaper distributor
- Travel agent and transportation

## **10. Wholesale, retail trade & repair service sector**

The major enterprise activities included in these sectors are:

- Alcohol and soft drink distributor
- Chat, Coffee production and cattle export
- Drug and Medical Supply Whole Sale
- Oil Cereals Export
- Export of garment and textile products and Export of Leather Products & Incense
- Vehicle Spare part & Maintenance
- Trade center, Super market and shopping center
- Small, Medium & heavy vehicle Renovation and maintenance

### **3.5. Knowledge Representation**

The Acquired knowledge is organized so that it will be ready for use. Knowledge representation and it is the systematic means of encoding knowledge of the human expert in an appropriate

medium like predicate calculus and logic, rules, semantic networks, frames and structured objects [49].

The acquired cases are represented using one of the different case base representation methods that are appropriate for the researcher. Case base representation methods include feature-value case representation, relational database case representation; predicate based case representation and soft computing case representation methods.

For this research feature-value case base representation method is used. Because, representing cases using feature-value case based representation is that this approach uses old experiences to understand and solve new problems. It also reuses its solutions and lessons learned in the past for future use and it represents cases in an easy way by using attribute and value pair representation [50], [51]. The algorithms used to calculate the similarity of cases in a case base representation is nearest neighbor retrieval algorithm. The similarity function of nearest neighbor retrieval algorithm involves in computing the similarity between the stored cases in the case base and the new query. After that, selects the most similar stored cases to the query.

### **3.6. Case structure for enterprise sector and enterprise activity selection**

For this research the case structure has two important parts. The first one is the problem descriptions or situation and the second one is the solution.

**Problem Description/Situation:** It is the part of the case structure that is consisted of attributes which describes the problem to be solved.

**Solution:** This part of the case structure provides the recommendation. (Enterprise sector and activity) based on the given members information.

Therefore, the researcher identifies different description and solution attributes with the help of enterprise experts and from the previous members cases recorded data set. But, there were different challenges during identification and representation of case structure. For instance, the first challenge was the recorded number of member's cases was too many for case base development. Due to this the researcher selected 5133 correctly fill successful previous members cases from 50,000 successful previous members cases and select 11 description and 3 solution attributes from total of 16 attribute to the development of case based recommender System. The

most important attributes that affects the selection of enterprise sector and activities decisions are listed as follows:

Attribute name	Parameter
Age	Description
Gender	Description
Capital	Description
Educational Level	Description
Form of ownership	Description
No. of members	Description
Region	Description
Zone	Description
Woreda/town	Description
Enterprise type	Description
Interested Enterprise activity	Description
Recommended Enterprise sector	Solution
Recommended Enterprise activity	Solution
Explanation	Solution

Table 3.2: The Case Structure for enterprise sector selection

The following short descriptions are used to explain the selected attributes:

**Age:** is the age of members. If members are engaged in a partnership or joint, the value of age is assigned plc instead of number.

**Gender:** is the sex of members as male and female. In this case the values of this attribute are male or female or plc. If members are formed in either partnership or joint, the value of gender is plc instead of male or female.

**Capital:** is the total amount of capital needed to start the selected enterprise sector.

**Educational Level:** it indicates the grade levels as well as the experience of members before starting the enterprise selection. Additionally it has also Plc value because assigning all the members values is difficult, if they are formed in a group.

**Form of ownership:** this attribute contains: the members form of ownership like private, cooperation (It is also an enterprise owned by a group of persons who take full part in the activity of the enterprise by coordinating their knowledge and assets) and trade association (Provide representatives and other collective products and services to businesses with a common interest).

**Number of members:** it contains the number of members initially and currently as well as it contains the temporary and permanent employees.

**Region:** is the regional or city administration place where members are planned to put their enterprises. The value of this attribute is Addis Ababa, Afar, Amhara, Benishangel gumze, Dire Dawa, Harari, Gambella, Oromia, SSNPR, Somali and Tigray.

**Zone:** is the zonal location of the enterprise. Since each region has its own zone, members selects specific zones based on the selected region.

**Woreda/town:** is the woreda or town location of enterprise that specifically facilitates service to members.

**Enterprise type:** contains values small and micro enterprises, which is categorized based on the amount of capital required.

**Member's interested enterprise activity:** is the process of member's interested activity that wants to start. So members fill their query as interested sector.

**Recommended enterprise sector:** it is the solution and provides a recommended enterprise sector based on the similarity of cases.

**Recommended enterprise activity:** it is put under a solution and provides a recommendation enterprise activity.

**Explanation:** is used to give explanation and description about the recommended enterprise activities.

## CHAPTER FOUR

### 4. DESIGN AND IMPLEMENTATION OF THE PROTOTYPE

This chapter deals with the design and implementation of the prototype recommender system for the selection of enterprise sectors and activities. After collecting the previous members cases' from FeMSEDA domain experts, members and document analysis, the next task is designing the prototype of recommender system for the selection of enterprise sectors by using the tool called jCOLIBR 1.1, it is a CBR framework used to design prototype for recommender system. For this research nearest neighbor retrieval algorithm is used. This is because jCOLIBRI uses this algorithm for retrieval task. Nearest neighbor algorithm retrieves the case which is nearest to the user's query by measuring its similarity with the cases. Nearest neighbor retrieval algorithm is also suitable when there are attributes which have numeric (continuous) value.

#### 4.1. Designing the architecture of case based recommender system

Figure 4.1 shows the architecture of the prototype for the selection of enterprise sectors and activities. The researcher collects enterprise members' cases from domain experts, members and document analyses then several preprocessing tasks were performed such as: select attributes, assign attributes, configure connectors etc. After performing all the tasks it is stored as a general knowledge or case base in order to build the case base recommender system.

As the new query/problem is entered, the prototype of the system matches the new case to the solved case in the case base of the system by using similarity measurement. When relevant cases are found within the case base, then the prototype system ranks the relevant retrieved cases based on their local similarity. Next, it proposes a solution.

The similarity between query and existing case is approximate, the proposed solution needs modification (adoption of solution) to fit the new case/query. At the end, the best modified solution should be stored into the case base for future use. The case base updates incrementally when the system learns from new case used by the members.

The proposed solution can be derived directly from a retrieved case that matches exactly or partially to the problem of the new case. Partial match of retrieved cases means some attribute values of the existing case and new cases/query are the same and some attribute values are

different. Using the proposed solutions directly may have a risk because some attribute values need of editing/changing based on different conditions. As a result the user of the system has made an adaptation on the proposed solution having differences between the proposed case and the new case. Additionally, case contradictions are revised because there are situations where previous cases attribute values are not similar with the new case/query attribute values. There is no similarity between the existing case and new case means there are no previous stored cases having similarity with the new case/query in all attribute values. Therefore if there is no similarity between the existing and new case, the proposed solution cannot give recommendation to new cases. In such cases, this new case or problem can be revised and stored in the case base. Finally, the revised solution or stored cases is retained in the case base for problem solving in the next time.

### FeMSEDA

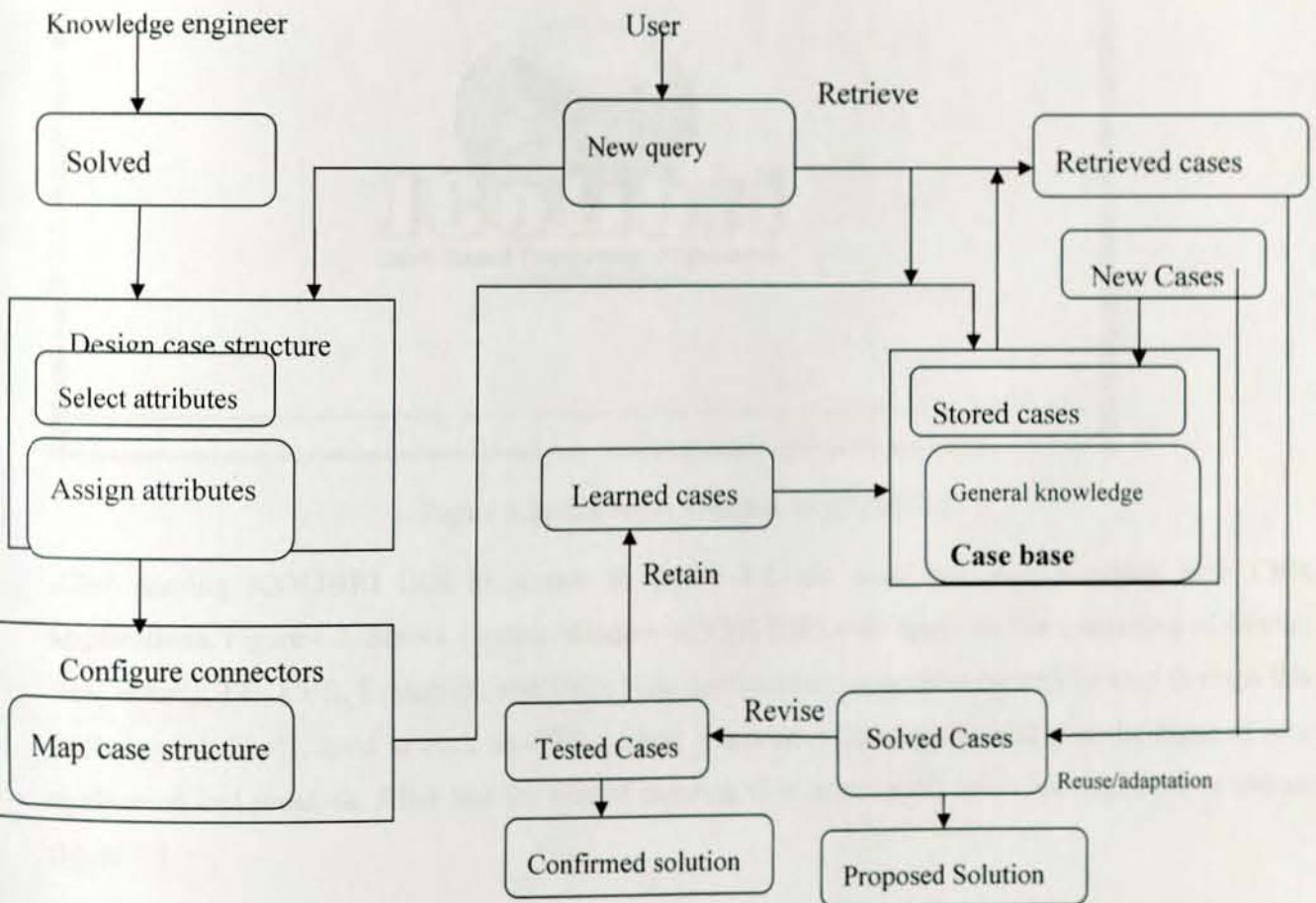


Figure 4.1: Architecture of CBRSESAEAS (source: [7])

As shown on figure 4.1: the tool jCOLIBRI follows the design to perform the implementation of this study. This figure is adopted from other researchers but it is modified in to this study context.

#### 4.2. Case based recommender system for the selection of enterprise sectors and activities

As presented in [52] jCOLIBRI has been constructed as a core module to offer the basic functionality for developing CBR application. jCOLIBRI can be started by clicking on exe file jCOLIBRIGUI.bat. GUI of jCOLIBRI helps us to create new CBR application with predefined task and methods. These predefined tasks and methods are in the files of **task.xml** and **methods.xml**. The main task of building CBR application is configuration where users define the task and assign methods for each task. So after clicking the jCOLIBRIGUI.bat file the window shown in figure 4.2 appears.

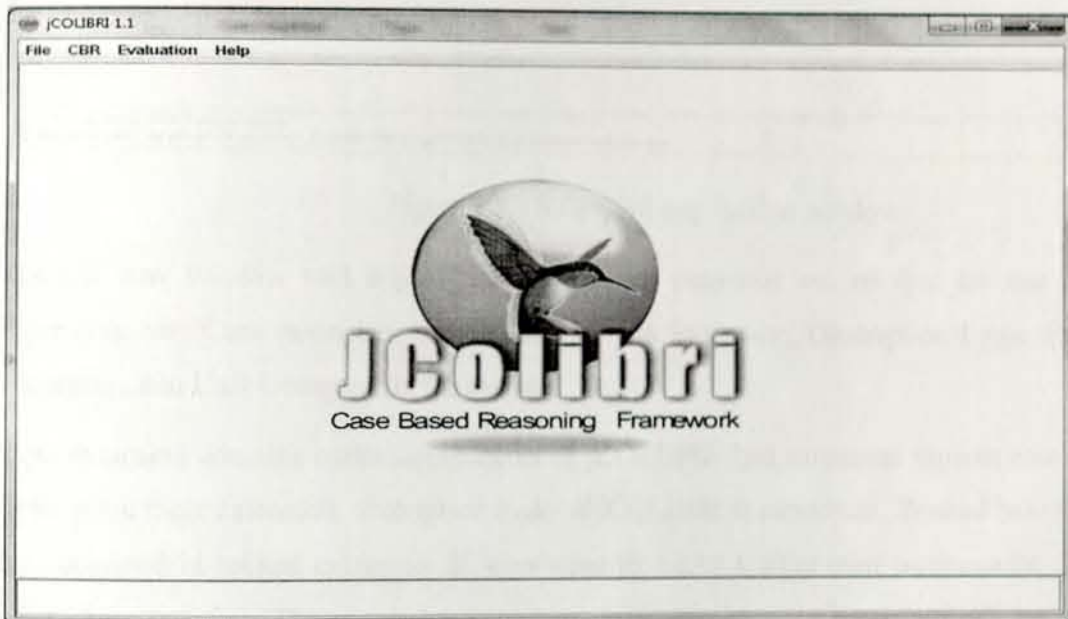


Figure 4.2: The Main Window of jCOLIBRI

After running jCOLIBRI GUI as shown in figure 4.2, the next task was **creating new CBR applications**. Figure 4.2, Shows the main window of jCOLIBRI with upper toolbar consisting of 4 menu lists, namely, file, CBR, Evaluation and Help. New application can be develop step by step through this GUI. For this firstly, need to click on CBR toolbar, select new CBR system and give the name of new application and press ok. After that the box of entering new application name are displayed as shown figure 4.3.

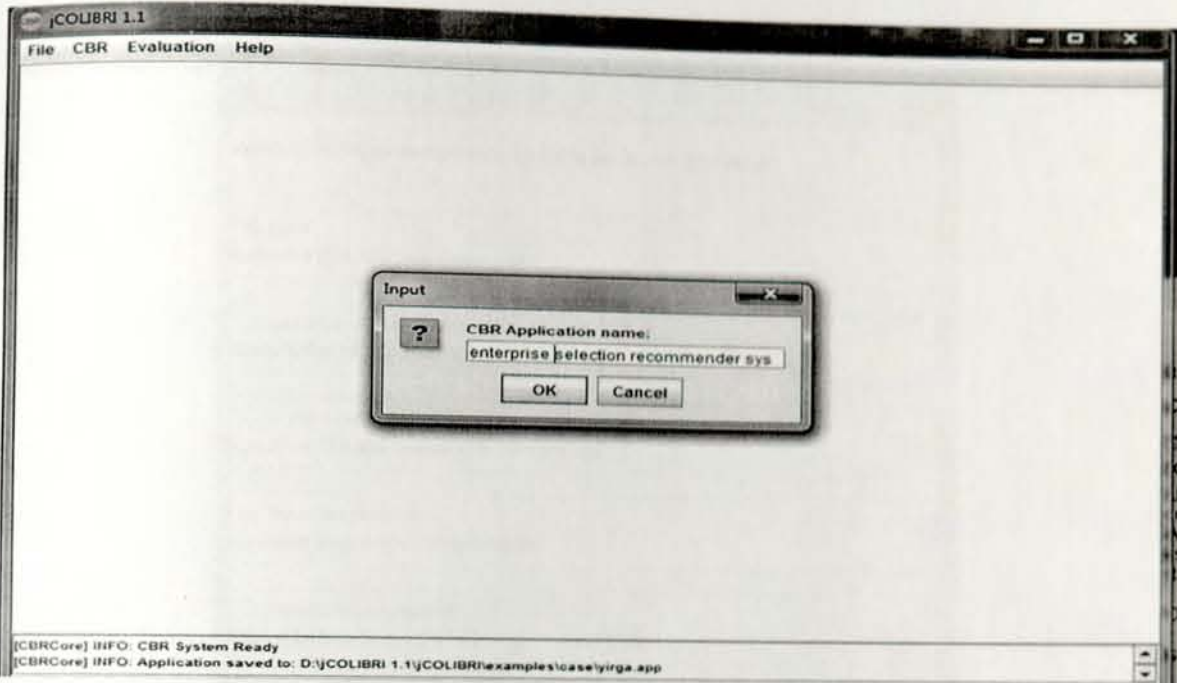


Figure 4.3: New CBR application window

Then, a new window will appear then select one extension out of five for our application. Five extensions are: Core extension, Case Retrieval Nets Extension, Description Logic Extension, Textual Extension, and User Components Extension.

Core extension contains basics components of jCOLIBRI. 2nd extension support case retrieval nets. In description logic extension, description logic of jCOLIBRI is supported. Textual base CBR components are supported in textual extension. If, user want to define his/her own components. Then, user has to select 4th extension. The researcher uses core extension because it contains all basic components of jCOLIBRI that are needed to make case based recommender system for this research.

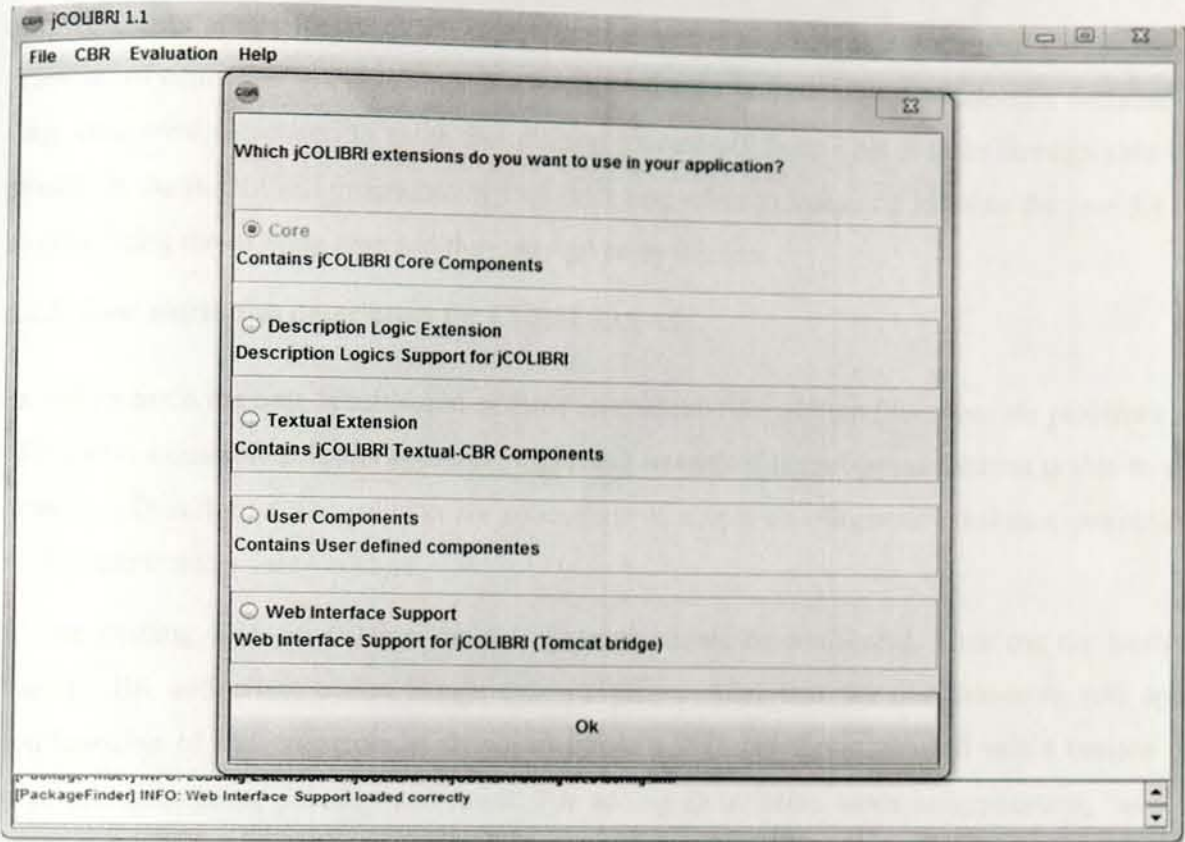


Figure 4.4: Types of jCOLIBRI extensions

#### 4.2.1. Building the Case Base

After collecting all successful enterprise members' cases from FeMSEDA, all the acquired cases are stored as plaintext file in a feature-value representation format. Feature value representation means each attribute has its own value in a column and row format.

The case base is presented as a plain text comprising of  $N$  columns representing case attributes ( $A_1, A_2, A_3, \dots, A_N$ ) and each  $M$  rows representing individual cases  $C$  ( $\{C_1, C_2, C_3, \dots, C_M\}$ ) each attribute has a sequence of possible values associated to each column attribute  $A = \{V_1, V_2, V_3, \dots, V_k\}$ .

#### 4.2.2. Case Representation

The case representation has been formulated in the way that easily represented in jCOLIBRI. Designing of such case structure helps to define the features available in the cases and used to measure the similarity between existing cases and the new case (query). The general application of this research is to

retrieve similar cases to the query from the case base that can guide enterprise members', solving problems of confusion and transforming a recommendation in their enterprise selection decision process. Case base were structured to make the retrieval process efficient. This is done through case indexing process in the jCOLIBRI programming tool. Indexing refers to assigning index to the case for retrieval by comparing the existing case and the query given by the user.

### **4.2.3. Case attributes description for CBRSESAEAS**

In this research the case is composed of three components: description (describes the problem), solution (represents a possible solution approach) and result (reveals if the proposed solution is able to solve the problem). Description and solution are collections of simple or compound attributes, permitting us to build a hierarchical case structure.

Before creating CBR application the case structure should be configured. Then use the toolbar menu button CBR and select option mange case structures. After that the new windows will appear for configuration of case structure as shown in figure 4.5. In the figure, the left side's options are case structure description, solution and result. For adding Description cases in application, "add simple" button is used, which is used to add simple attributes only. "Remove button" is used to remove the description attribute if it's not necessary. When select one of description attributes, for instance select on type of enterprise, and then its properties will appear on the right side. Name, type, weight and similarity of case are properties of description attribute. "Apply changes" button is used to change the properties of attribute description. Local similarity is used for computing the similarity of each attributes. After defining the structure of cases, case structure is saved in xml file. During configuration of case structures, jCOLIBRI creates codes automatically and saves in xml file format. Then, by applying the above all steps the following window were appeared.

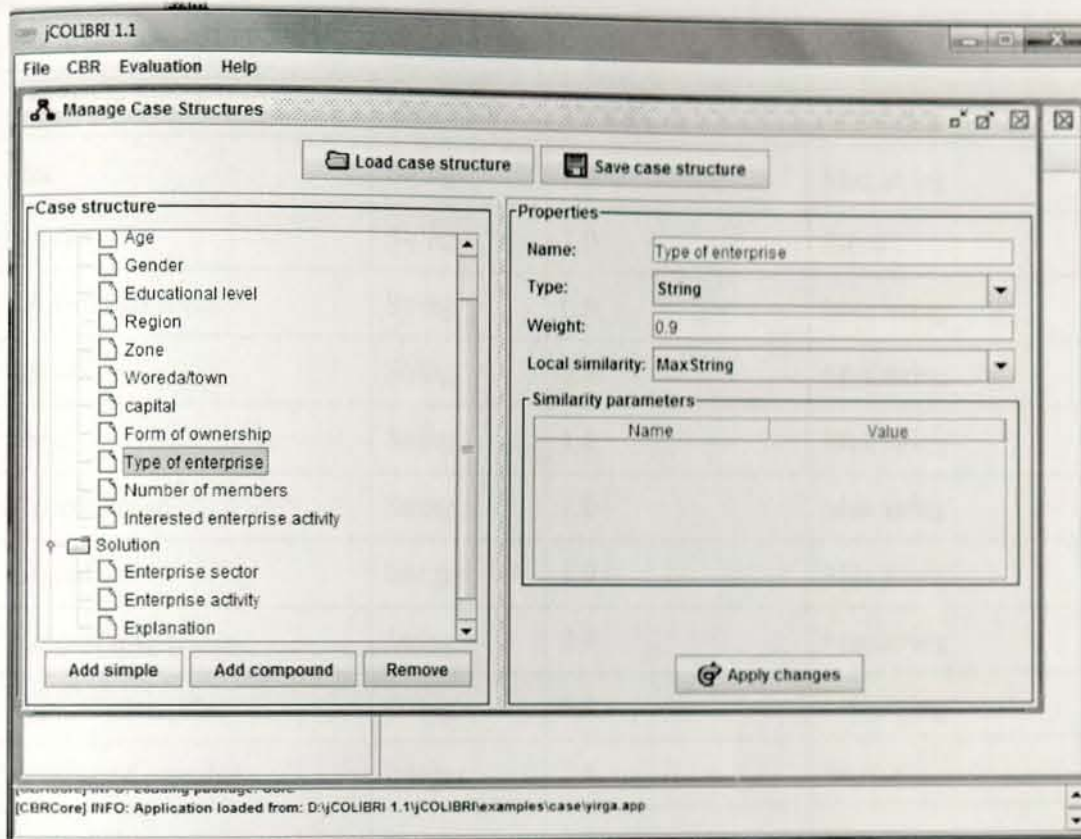


Figure 4.5: Defining case structure

After defining the case structure, members input the value of all description attributes and measuring the similarity between the existing cases attributes value and new cases attribute value. For this research the solution attributes include recommend enterprise sectors, recommended enterprise activity, and the explanation facility about the recommended enterprise activity.

The following table 4.1 shows the description of case attributes and solution attribute regarding name, data type, weights, local and global similarity.

<b>Significant attributes</b>			
<b>Attribute name</b>	<b>Data type</b>	<b>Weight</b>	<b>Local similarity</b>
Age	String	1.0	Max string
Gender	String	1.0	Equal
Educational level	String	0.9	Max string
Region	String	0.9	Max string
Zone	String	1.0	Max string
Woreda/town	String	1.0	Max string
Capital	Integer	1.0	Max string
Form of ownership	String	0.9	Max string
Type of enterprise	String	0.9	Max string
Number of members	Integer	0.8	Interval
Interested enterprise activity	String	1.0	Max string
<b>Solution</b>			
Enterprise sector	String	1.0	Equal
Enterprise activity	String	1.0	Equal
Explanation	String	1.0	Equal

Table 4.1: Descriptions and Weight of the Selected Attributes

As shown in the above table these attributes are the most significant for the selection of enterprise sector activities and attributes have their own different values which indicates that having high weight have most significant. Based on this, educational level, region, form of ownership, type of enterprise, number of members have 0.9, 0.9, 0.9, 0.9, 0.8 weights respectively. The weight value of each attribute has been assigned by using information gain attribute selection algorithm and domain experts. The local similarity of most description attributes is maximum string. This is because the similarity between query and cases

can be calculated with maximum string length. Few attributes such as gender have equal similarity weight because local similarity needs exact match of existing cases and new case (query).

After identifying relevant attributes of the case, the next task is definition of appropriate similarity measure in jCOLIBRI. jCOLIBRI follows both local and global similarity measures. Different types of local and global similarities are used in this research.

**Local similarity:** Local similarity measure divides the similarity definition into a set of local similarity of each attribute. Three types of local similarity measurement:

**A. Equal:** The input query and cases in the case base must match to get the result; if there is no match between input query and cases, matching will fail.

**B. Interval:** When we select similarity interval and adjust interval value, jCOLIBRI matches value keeping in mind that interval. Exact value match is not compulsory in interval local similarity.

**C. Max string:** if we select the max string local similarity, the system matches by using the maximum string length.

**Global Similarity** is linked with compound attributes and used to get similarity of collected attributes in unique similarity value. Global similarity calculates the final similarity measure.

**A. Average:** - it is a type of global similarity that considers the average of all attribute similarity values.

#### 4.2.4. Managing connectors

Once case structures are configured in jCOLIBRI, recommender systems must access the stored cases from case base. jCOLIBRI supports both SQL database and plain text file to store its cases base.

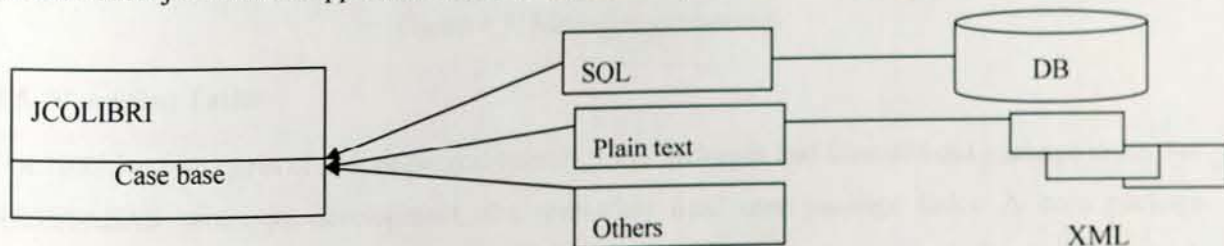


Figure 4.6: jCOLIBRI case base schemas (source: [52], [53])

In this research plain text connector is used as a case base storage. The connector maps the case structure to its column from plain text file which is saved in .txt file format and later saved as XML file like that of case structure. One of the most important tasks in managing connectors is specifying the correct path of case structure and file path. The case structure path is used to access and match attributes from case structure and file path is used to specify the .txt file that contains the case base. Delimiter of this connector uses comma (,) to separate value of each attribute in the case.

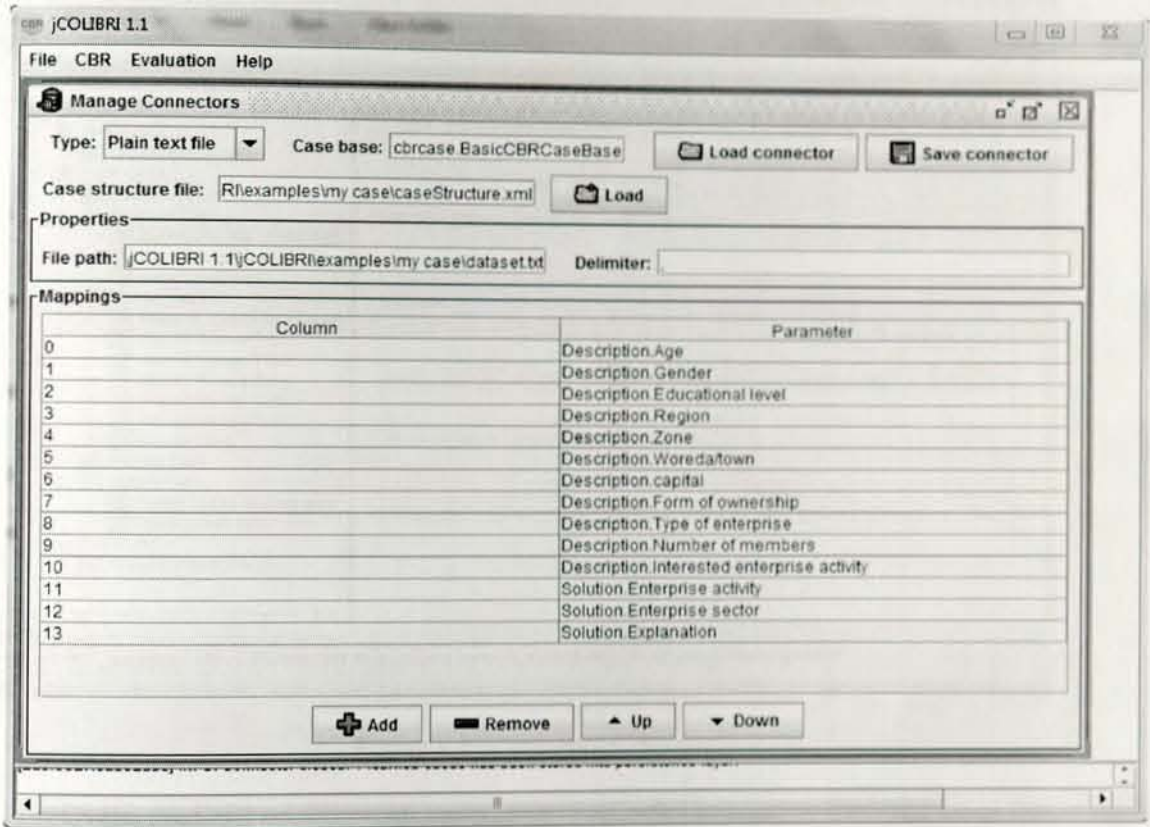


Figure 4.7: Managing connectors

#### 4.2.5. Managing Tasks

jCOLIBRI has two types of task packages, namely, Core packages and User defined package tasks. For CBRSEAEAS prototype development, the researcher used core package tasks. A core package contains all classes that represent core functionality of a CBR application such as the domain model, case bases, similarity functions and retrieval algorithms. Core packages also have predefined tasks and methods that used to configure new system by reusing the tasks rather than using tasks or methods

defined by the system developer itself like user defined packages. A core package in jCOLIBRI contains: main CBR cycle (retrieve, reuse, revise and retain) and PostCycle. So these core package components are used to configure all the necessary case base and connectors for the development of CBRSEAEAS prototype.

The left side of figure 4.8 shows PreCycle, CBR Cycle and PostCycle.

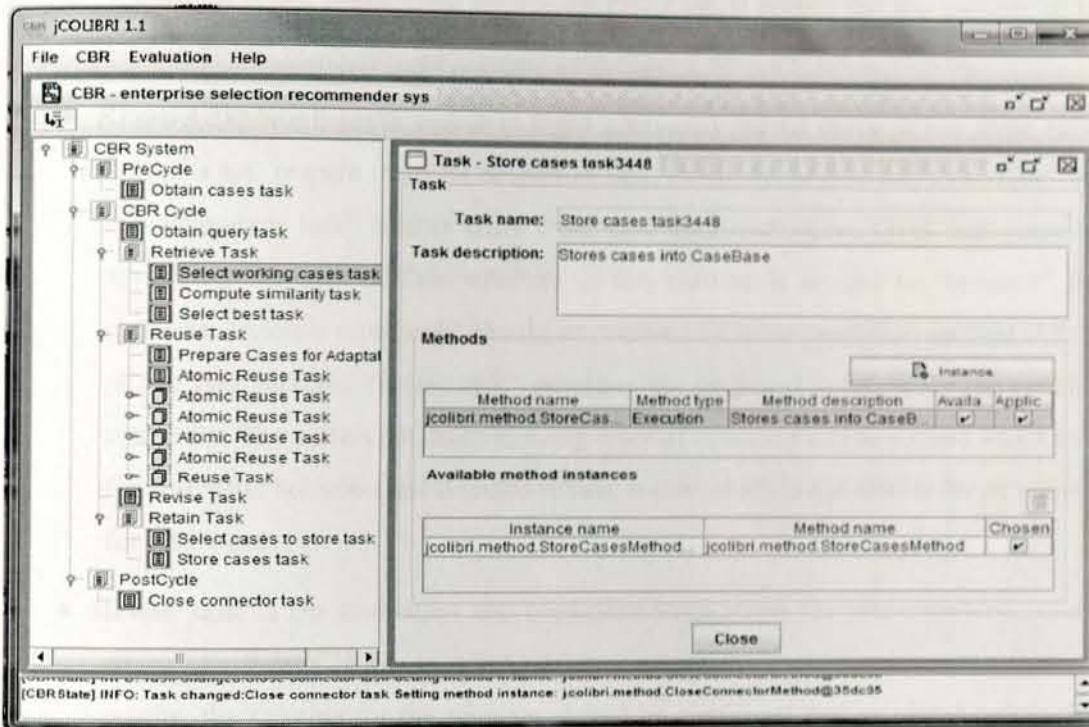
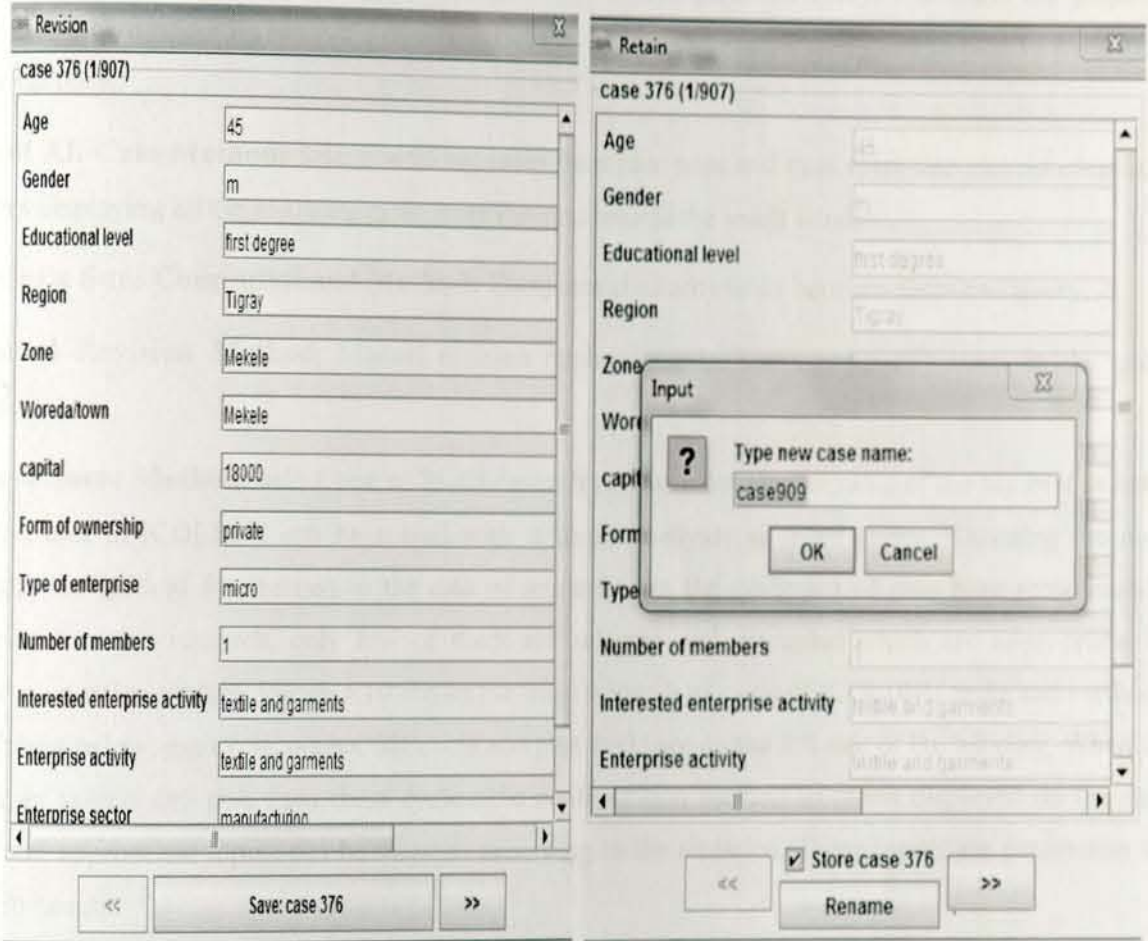


Figure 4.8: Configure CBR application

- **PreCycle**, the CBR configuration starts from “PreCycle” in order to loads the cases from case base. This precycle have only one sub task called “obtain cases task” which is used to load the cases from the case base by defining the path of the connector. And make instance to tasks and methods.
- **Main CBR cycle** is the main task of CBR cycle and it also has sub tasks. The developer has to give path of case structure that is saved in xml format in Main CBR cycle sub task called “obtain query task”. “Obtain query task” is used to knows the number members cases attributes that are available after the path is assigned. In addition to obtaining query task, main CBR cycle contains: retrieve tasks, reuse tasks, revise task and retain tasks.

- **Retrieve tasks** is used to retrieve cases from the stored case base. Retrieve tasks have also subtasks include select working cases task, compute similarity task and select the best case. “Select working case task” selects cases from case base and stores them into current context. “Compute similarity task” compute similarity of the stored cases with the case entered by the user using the query window. “Select best case” shows the best matched case after computing the similarity of stored cases against the new case. It means that the number of best matched case is shown to the user depending on the method used and the threshold.
- **Reuse/Adaptation tasks** enable to reuse previously stored cases. It has three subtasks. These subtasks are: prepare cases for adaptation task, atomic reuse task and reuse task. “Prepare cases for adaptation task” selects cases from case base and stores them into context. Here also specifying the path of case structure in this method is needed to “instance” the tasks and methods. “Atomic reuse task” should be resolved by reuse resolution method. After the process of the two subtasks “Reuse task” generates the proposed solution for the problem based on similarity. The system can learn at every entry of new case and new users adopt this knowledge for enterprise selection and it reuses if there a case which is not similar for new query and stores for future use.
- **Revise task** is the evaluation and correction stage about the recommended solution in reuse phase. As shown in figure 4.9(A) after selecting the most similar cases from the retrieved results, the solution for the problem should be confirmed and validated before the solution is stored for future use.
- **Retain tasks** are also used to CBR case retention on a persistence layer. It has also its own subtasks like “select cases to store task” and “store cases task. “Store case “was used to type a new case name as shown in figure 4.9 (B). Select cases to store task give authentication to the user for storing case. The store cases task enables to store cases into the case base. Retain task is performed after having confirmation in revision phase. So after the evaluation and correction of retrieved cases in revise task the problem together with its solution will be stored in case base.

- **Post Cycle:** is the last task in managing tasks in jCOLIBRI. Post Cycle task have only one sub task called “close connectors task” which is usually executed after the main CBR cycle. Its main task is to close a connection between case base and GUI.



(A)

(B)

Figure 4.9: Revision and retain tasks

#### 4.2.6. Managing methods

The method Computes similarity tasks between cases and query packages store classes that resolve the task. These classes can resolve the CBR cycle using programming or using graphical user interface (GUI). All tasks in jCOLIBRI should have their own methods to be assigned in order to achieve its recommendation goal.

**Load CaseBase Method:** This method returns the whole available cases from the case base to designer and use connector as parameter to retrieve case base.

**Configure Query Method:** This method obtains and configures the query. It displays the graphical user interface window by receiving case structures as an input to request query and to receive cases from the case base.

**Select All Case Method:** Selects working cases from case base and store them into current context. It allows displaying all the available cases from the case base to the result window.

**Numeric Sum Computational Method:** Computes similarity tasks between cases and query.

**Manual Revision Method:** Manual revision method enables users to modify cases in the query window.

**Select Some Method:** select best of found cases. It returns most similar value of the top best selected case. Tasks in jCOLIBRI can be solved with different methods as listed above. Choosing the most appropriate method for the task is the role of researcher in the designing of case base recommender system. For this research, only few of them are selected and discussed which are appropriate for recommendation system. Figure 4.10 shows the main window of cycle of jCOLIBRI tasks and methods. As shown below, pre cycle, main CBR cycle and post cycle are on the left side of the window. When the designer selects any task from these cycles, the configuration method windows displayed on the right side and appropriate inputs can be selected according to the situation. These inputs are parameters for new instances.

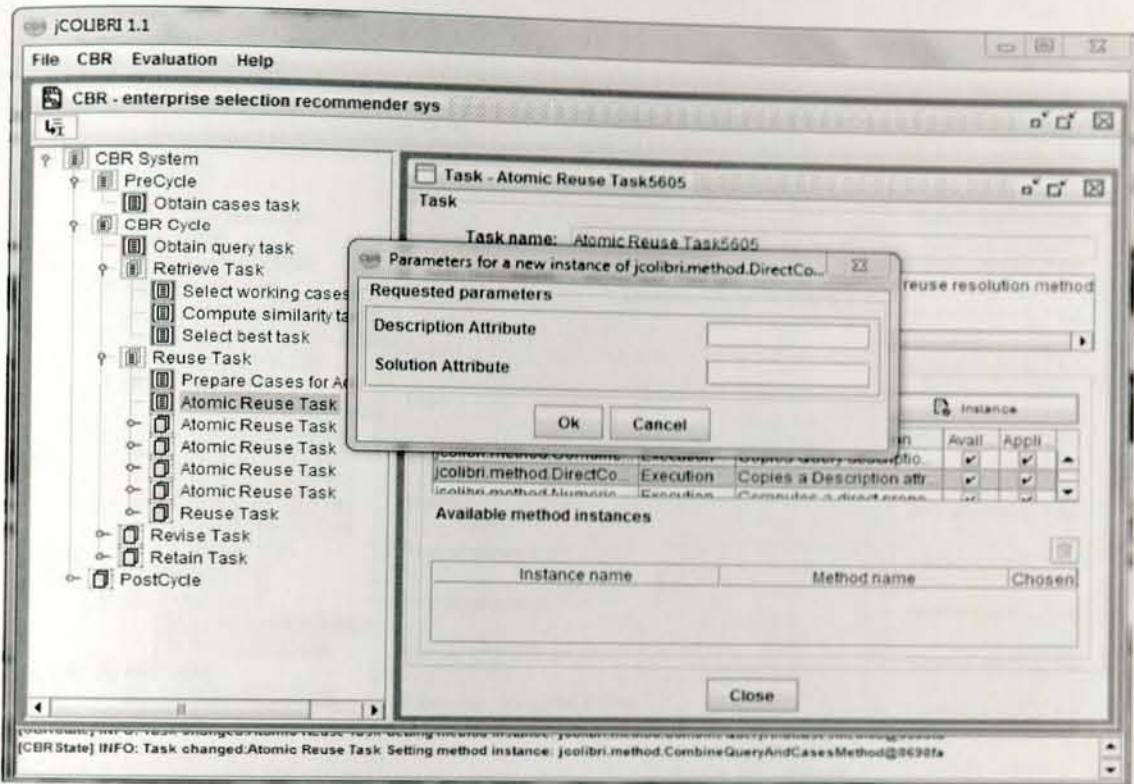


Figure 4.10: Configuration of tasks and methods

#### 4.2.7. Deploy the case base recommender system

After defining and configuring all the necessary steps required designing case base recommender system in jCOLIBRI, testing the recommender system application is the next step as shown in figure 4.11.

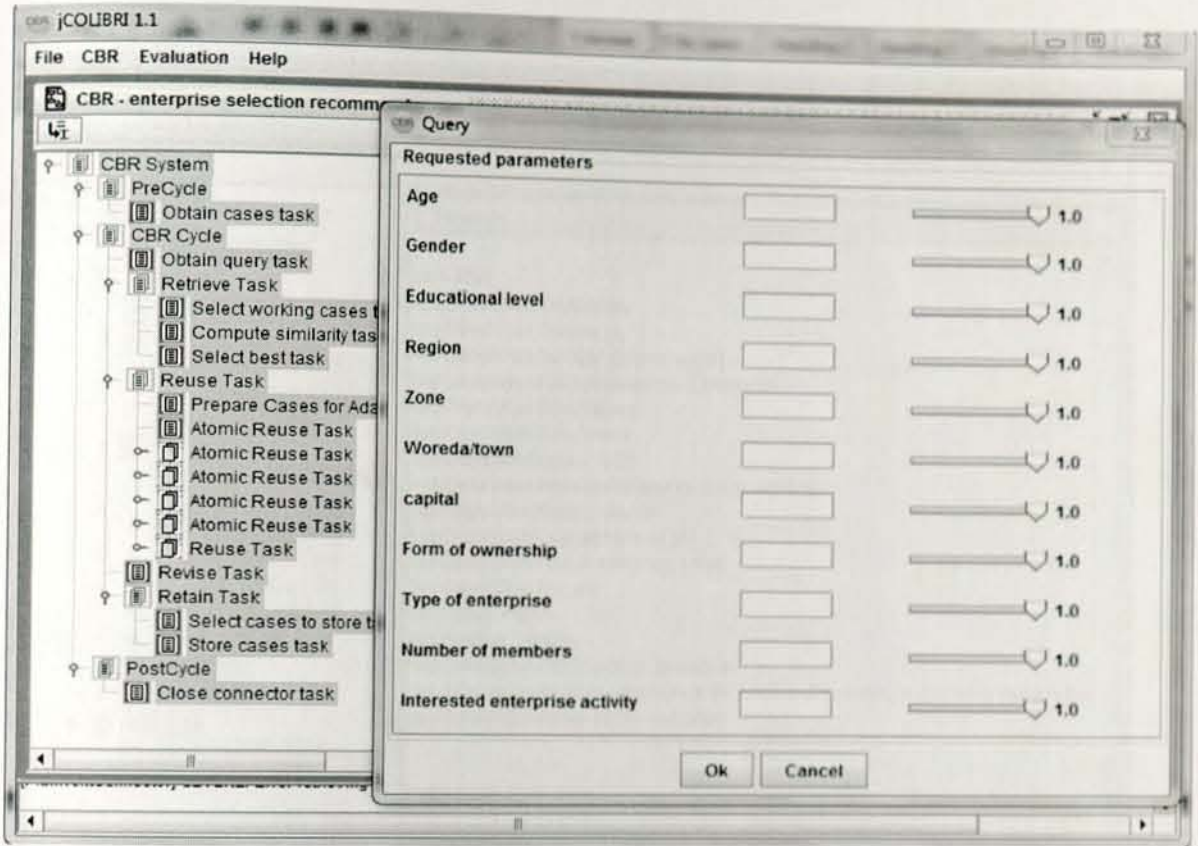
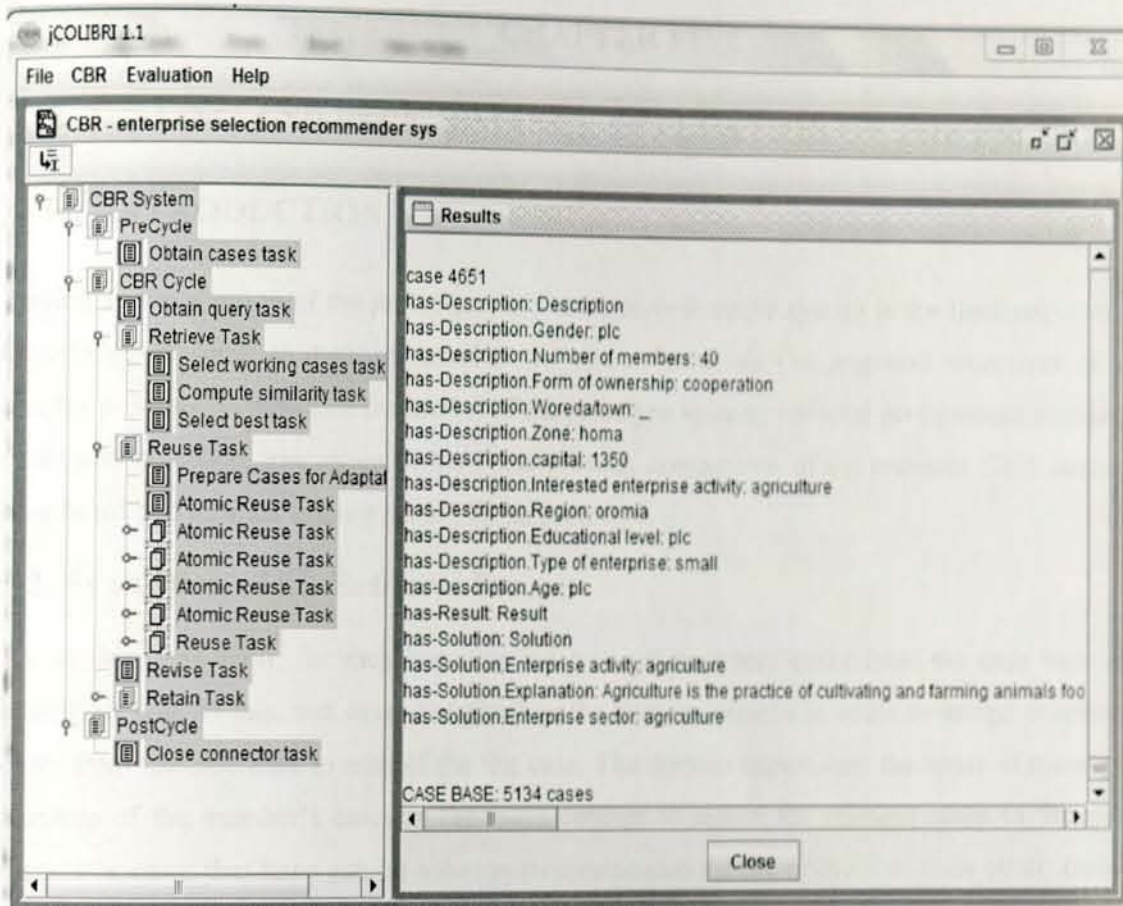


Figure 4.11: Query entry window in to the case base

#### 4.2.8. Explanation Facilities

The explanation facility in this study is used to give explanation about the recommended enterprise activity given by the system. As shown in figure 4.12, once the system reaches final decision on the recommendation, it gives explanation about the selected enterprise activity.

The system gives the name of selected enterprise activity which is agriculture, then a brief explanation about agriculture, then in which sector agriculture is categorized and finally all the necessary descriptions are there.



**Results**

```

case 4651
has-Description: Description
has-Description.Gender: plc
has-Description.Number of members: 40
has-Description.Form of ownership: cooperation
has-Description.Woreda/town: _
has-Description.Zone: homa
has-Description.capital: 1350
has-Description.Interested enterprise activity: agriculture
has-Description.Region: oromia
has-Description.Educational level: plc
has-Description.Type of enterprise: small
has-Description.Age: plc
has-Result: Result
has-Solution: Solution
has-Solution.Enterprise activity: agriculture
has-Solution.Explanation: Agriculture is the practice of cultivating and farming animals for
has-Solution.Enterprise sector: agriculture
  
```

CASE BASE: 5134 cases

Close

Figure 4.12: Explanation facility for enterprise activity

## CHAPTER FIVE

### 5. TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE

#### 5.1. INTRODUCTION

Testing and evaluation of the prototype case based recommender system is the final step that helps the knowledge engineer to measure whether the system achieves the proposed objectives or not. This chapter presents performance evaluation of the prototype system, retrieval performance evaluation using recall and precision, evaluation of the reuse process, comparison of the previous CBR cases with this case based and user acceptance testing of the prototype.

#### 5.2. Evaluation of the Retrieval Process

To do this evaluation, for each test case the relevant members cases from the case base should be identified. Due to this, test cases are given to the domain experts in order to assign possible relevant cases from the case base to each of the test case. The domain expert uses the value of recommendation attribute of the member's cases as the main concept to assign the relevant cases to the queries, i.e. member's cases that have similar solution (recommendation) are relevant to each other. Based on this concept recall and precision are calculated.

Table 5.1 below shows sample test case with their corresponding relevant member's cases that are assigned by the domain expert from the case base.

Test case	Relevant cases from the case base
Case 5047	Case347, Case 5049, Case4010, Case2609, Case123, Case4013, Case4008, Case255
Case 4881	Case40, Case303, Case4437, Case2288, Case22, Case2292, Case4434, Case38, Case4439
Case 4081	Case486, Case4, Case4083, Case1, Case88, Case4084, Case4417, Case5, Case22, Case 3053
Case 10	Case 2, Case 4, Case 3053, Case 1, Case 4081, Case 4086, Case 5, Case 3473, Case 4425
Case 1	Case 4, Case 4081, Case 4086, Case 5, Case 7, Case 2, Case 3053, Case 10
Case 4884	Case885, Case 4886, Case 87, Case 488, Case 4445, Case33, Case 3775, Case 76, Case 4881
Case 823	Case 82, Case 828, Case 830, Case 26, Case 824, Case 89, Case 75
Case 2290	Case287,Case2291, Case92, Case2293, Case88, Case 289, Case 2294, Case 4436, Case 4446
Case 4870	Case 473, Case 71, Case 4077, Case 4088, Case 4069, Case 404, Case 4072, Case 407
Case 3057	Case 3051, Case 53, Case 4083, Case 84, Case 3055, Case 306, Case 73

Table 5.1: relevant cases assigned by domain experts for the sample test case

After the identification of the relevant cases to the test cases by the domain expert after the system is developed, precision and recall values were calculated with threshold interval to test the prototype using the test case query to know the performance of the system. There is no standard threshold for degree of similarity that has been used for retrieving relevant cases. That is why different researchers use different case similarity threshold to measure the performance of their system. For this research 0.7 and 1.0 threshold values are used.

The researcher conducting two hundred thirteen (213) sample experiments to measure recall and precision and after similarity computation, the first 7(seven) best cases are retrieved assuming that the first 13 relevant cases are retrieved from the case base as shown in the above table ( table 5.1).

<b>Test cases</b>	<b>Recall</b>	<b>Precision</b>
Case 5047	0.87	0.61
Case 4881	0.78	0.69
Case 4081	0.70	0.76
Case 10	0.78	0.69
Case 1	0.87	0.61
Case 4884	0.87	0.69
Case 823	1.0	0.53
Case 2290	0.87	0.69
Case 4870	0.87	0.61
Case 3057	1.0	0.53
<b>Average</b>	<b>0.86</b>	<b>0.64</b>

Table 5.2: recall and precision results for the sample test case

As shown in table 5.2 above, recall of each test case can be calculated by dividing the number of relevant retrieved cases with total relevant cases. For instance, for case5047 seven relevant cases are retrieved out of the total eight relevant cases in the case base. When we calculate its recall value, the result is 0.87 (87%). When we come to the precision of the system, it is calculated as the number of relevant cases retrieved divided by total retrieved cases. For instance, case5047 contains 8 relevant cases and total of 13 retrieved cases. Therefore the precision value of case5047 is 0.61 (61%). Other all the precision and recall values are calculated with similar way.

Generally, the average recall and precision results 86% and 64% respectively which is also a promising result. As seen in the table 5.2, for every test case more than average is registered both recall and precision. In terms of recall this research achieved a very good result than precision is somewhat lower compared to the average recall.

### 5.3. Evaluation of the Reuse Process

The goal of reuse process in this research is to evaluate the proposed system whether recommend enterprise sectors and enterprise activity correctly for new member's cases, i.e. to solve the problem correctly. The performance of the reuse process is measured by using accuracy. Accuracy is one of the useful measurements in case based reasoning. Accuracy is defined by the percentage of the number of correctly recommended cases divides by the total sample case [54]. This measurement had been used by [55] on their research.

The result in table 5.3 shows that the reuse process also registers above average, which is a good result.

Total number of tested case query	Total number of correctly recommended cases	Accuracy
213	191	90%

Table 5.3: Accuracy value of the reuse process

### 5.4. Comparison of the Performance of CBRSESAEAS with Previous CBR Systems

The performance of the system is compared with the previously conducted researches. The previous research focus on retrieval phases of the case based reasoning system. Thus, the recall and precision value of retrieval performance of the systems developed by the earlier thesis research are compared as shown in table 5.4 below.

Domain area and researcher	Programming tools used	Retrieval task		Reuse task
		Recall	Precision	Accuracy
AIDS Alemu (2010)	jCOLIBRI	72%	63%	Not evaluated
Hypertension Henok(2011)	Python	86.1%	60%	88.89%
Mental Health Getachew (2012)	jCOLIBRI	82%	71%	Not evaluated
Field of study selection Biazen (2013)	jCOLIBRI	85%	55%	Not evaluated
Investment yibeltal (2013)	jCOLIBRI	85%	64%	87%
Enterprise selection (2014)	jCOLIBRI	86%	64%	90%

Table 5.4: A comparison of CBRSESAEAS system with the previous CBR systems

As shown in table 5.4 the result of this study is better than others because, it uses many cases for case development. This leads the system to provide more case solution than the other developed systems for users query in this area.

### 5.5. User Acceptance Testing

During testing the user's acceptance, the applicability of the prototype is evaluated by potential users of the system. The FeMSEDA six (6) domain experts and ten (10) members from each enterprise sector which were actively participated from the beginning to the end of the system. During testing experts are requested to rank each parameter from to excellent by assigning value for poor=1, fair=2, good=3, very good=4, excellent= 5.

No.	Evaluation criteria	Performance value					Average	Percentage
		1	2	3	4	5		
1	Is the system is efficient in time			1	2	3	4.3	86
2	Easy to use			2	2	2	4	80

3	Is the interface is interactive for users			1	2	3	4.3	86
4	Relevancy of the retrieved cases for decision making				3	3	4.5	90
5	Adequacy and clarity of decision support				4	2	4.3	86
6	Relevancy of the attributes in representing members case				2	4	4.7	94
7	Is the explanation facility is clear				2	4	4.7	94
8	Significance of the system in the domain area				1	5	4	80
				Total average			4.3	87%

Table 5.5: The CBRSESAEAS system performance evaluation by the domain expert

As shown in the above table 5.5, regarding to the efficiency of the system 16% of the respondents said good, 34% respondents as very good and 50% of the respondents said as excellent. Similarly, use of the recommender system 33% of the respondents said good and 33% said very good and 34% said excellent. In the case of interface interactive, 16% of the respondents said good, 34% as very good and 50% as excellent. For the relevance of retrieved cases for decision making, 50% of the respondents said as very good and 50% of the respondents said excellent. The clarity and adequacy of decision support, 68% of the respondents said as very good and 32% of the respondents said excellent. For relevancy of the attributes, 32% of the respondents as very good and 68% and the clearness of explanation facility, 32% of the respondents as very good and 68% of the respondents as excellent. Finally for significance of the system in the domain area, 16% as very good and 84% of them said as excellent.

No	Evaluation criteria	Performance value					Average	Percentage
		1	2	3	4	5		
1	Is the system is efficient in time			1	4	5	3.4	68
2	Easy to use		2	2	6	2	4.4	88
3	Is the interface is interactive for users		1	4	3	2	3.6	72

4	Relevancy of the retrieved cases for decision making				5	5	4.5	90
5	Adequacy and clarity of decision support			2	4	4	4.2	88
6	Relevancy of the attributes in representing members case				2	8	4.8	96
7	Is the explanation facility is clear				3	7	4.7	94
8	Significance of the system in the domain area				2	8	4.8	96
				Total average			4.3	86.5%

Table 5.6: The CBRSESAEAS system performance evaluation by the enterprise members

As shown in the above table 5.6, for the efficiency of the system 10% of the respondents said good, 40% respondents as very good and 50% of the respondents said as excellent. Next, the use of the recommender system 20% of the respondents said fair, 20% of the respondents said good and 40% said very good and 20% said excellent. In the case of interface interactive, 10% of them said as fair, 40% of the respondents said good, 30% as very good and 20% as excellent. For the relevance of retrieved cases for decision making, 50% of the respondents said as very good and 50% of the respondents said excellent. The clarity and adequacy of decision support, 20% of the respondents said as fair, 40% of the respondents said as very good and 40% of the respondents said excellent. For relevancy of the attributes, 20% of the respondents as very good and 80% and the clearness of explanation facility, 30% of the respondents as very good and 70% of the respondents as excellent. Finally, for significance of the system in the domain area, 20% as very good and 80% of them said as excellent.

Based on eight close ended questions and five parameters both domain experts and members were evaluated after developing the system. The result shows above average (86.5% and 87%) to both members and experts respectively. This indicates users are satisfied.

## CHAPTER SIX

### 6. CONCLUSION AND RECOMMENDATIONS

#### 6.1. Conclusion

CBR is a part of AI which enables us to design an intelligent agent that makes decision from the past solved cases i.e. a new problem is solved by finding a similar past case and reusing it in the new problem situation. As compared to rule based reasoning, CBR can work with new cases that match partially to the case from the case base. However, rule based reasoning cannot solve a problem that doesn't exactly match with the rule of the system. This shows that rule based reasoning works in closed assumption where every fact are known and represented.

In Developing country like Ethiopia the advising system on MSEs remain at lower stage. Different factors affect the sector advising system in Ethiopia. These factors include lack of guide line or criteria to assign members in different enterprise sectors, shortage of skilled manpower in the area, lack of consistency of advising system, and lack awareness of members about the purpose of advising systems for the selection of enterprise sectors and activities.

The main goal of this research is to develop a prototype CBRSESAEAS system. The system aims to assist both the domain experts and new members in the processes of making proper enterprise sector and enterprise activity selection decisions from already solved cases. In order to, acquire knowledge domain experts, members and documents were used. Mainly previous member's cases were collected from FeMSEDA. The acquired knowledge was conceptually modeled using hierarchical structure conceptual modeling method. The Case representation method that is used in this study is feature value case representation method. Feature value case representation is applied to represent the knowledge before it has been codified using the jCOLIBRI tool. The prototype was developed by using jCOLIBRI 1.1 Programming tool.

Finally, the evaluation result shows that as retrieval performance of the prototype registers an average value of 86% recall and 64% precision, while its reuse performance registers an average value of 90% accuracy. The user acceptance evaluators (both domain experts and members) assign more than average

value for all parameters that are used in the user evaluation form for the prototype. So the average user acceptance evaluation achieved 87% and 86.5 % performance by domain experts and members respectively.

Furthermore, the following conclusions are drawn from the finding with regard to the research questions:

- The major attribute that have more influence in enterprise sector and activity selection are age, gender, capital required, educational level, enterprise location (such as region , zone, woreda), form of ownership, enterprise type, interested area of enterprise activity.
- The applicability of case based recommender system for enterprise sector and activity selection haven been proved with evaluations.
- The result of system performance indicated that users are satisfied with proposed system and the performance of the system validation result showed the system recommends highly acceptable enterprise sector and activities for members.
- In the proposed case based recommender system learning is made for new cases by incrementally updating in the existing case base for the purpose of future use as a case base.

## **6.2. Recommendation**

Even if, the results of this study are promising, there are problems that need further investigation for future work. Therefore, based on this study the researcher recommends the following issues as a future research direction:

- The attributes used for this research were collected from the previous members' cases from FeMSEDA. These attributes are not enough for the selection of enterprise sector and enterprise activity decision. So further research can be conducted by adding other important attributes such as land size required, marital status and level of risk taker for the enterprise sector and activity selection.
- To enhance the performance of the prototype case based recommender system of the system Hybrid strategy approaches can be investigated which combines rule based reasoning and case

based reasoning. Because, combining rule based and case based reasoning systems can avoid the limitations of each of them.

- The system was developed in English language and is difficult to understand by some members. Further investigation can be conducted by developing the CBRSESAEAS system in different local language. Because, this helps members to communicate using their own language with the case based recommender system.
- In this study the explanation facility given by the proposed system is not user interactive. Not user interactive means the explanation is given only once when the system assigns enterprise sector and activity selection to the new case. The explanation facility gives explanation only about the recommended (assigned) enterprise activity, but does not give explanation at any time the user needs explanation. So further research can be done to add explanation facility that can give explanations any time the user wants explanations in addition to explanation of the recommended enterprise activities.
- The developed system does not have any controlling mechanisms to control the dynamic nature of the environment. Environments change may affect the success and implementation of enterprise activities. So, future researches can be done on managing the dynamic nature of the environment to start enterprise activity. Further investigations can be done with case based reasoning system for other new sectors or organizations.

## REFERENCES

- [1] Gebrehiwot Ageba and Wolday Amha, "Micro and small enterprises (MSEs) Finance in Ethiopia: Empirical Evidence," volume 22, number 1, pp. 63-86, 2006.
- [2] Mulu Gebre Eyesus, "Growth of Micro Enterprises Empirical: Evidence from Ethiopia" Development Research Institute (EDRI), 2007.
- [3] Hk Hailu, "Success factors in Micro and Small Enterprises Cluster Development:"Case of Gullele Handloom Clusters in Ethiopia, Addis Ababa, Ethiopia, 2010.
- [4] Netsanet Jote, Cluster Approach for Enhancing the productivity and competitiveness of Micro and Small Enterprises /MSEs/ Case Study on Bamboo Micro and Small Enterprises, 2009.
- [5] *Micro and Small Enterprises Development Strategy*, Addis Ababa, Ethiopia, 1997.
- [6] Robin Burke, "Knowledge-Based Recommender Systems Department of Information and Computer Science," University of California, 1995.
- [7] Yibeltal Chanie, "Application of Case Based Recommender System in Investment Sector and Investment Activity Selection to New Investors:" In the case of Ethiopia, 2013.
- [8] Biazen Getnet, "Application of Case Based Recommender System to Advice Students in Field of Study Selection at Higher Education in Ethiopia," 2013.
- [9] Guy Shani and Asela Gunawardana, "Evaluating Recommendation Systems," 2010.
- [10] *Ethiopian Investment Agency, Manuals of Ethiopian investment guide*, Ethiopia, Addis Ababa, 2012.
- [11] Fabiana Lorenzi and Francesco Ricci, Case-Based Recommender Systems: a Unifying View, 2000.
- [12] Background Issue Paper on Legal Empowerment of the Poor Entrepreneurship, Addis Ababa, 2006.
- [13] *Ethiopian investment agency, Manuals of Ethiopian investment guide, Ethiopia*, Addis Ababa, 2013.
- [14] *Trade and development Office, Micro and Small Enterprises Development Strategy*, Ethiopia, Addis Ababa, 1997.
- [15] Getachew Wassie, "Application of Case-Based Reasoning for Anxiety Disorder Diagnosis, Addis Ababa, Ethiopia, 2012.
- [16] Ethiopia Tadesse Akana, "Application of Case-based Reasoning for Amharic Legal Precedent Retrieval:" A Case Study with the Ethiopian Labor Law, 2002.
- [17] *Federal Micro and Small Enterprises Development Agency, Manual on the Micro and Small*

- Enterprise in Ethiopia, 2013.
- [18] Endalkachew Mulugeta, "Underlying Causes of Micro and Small Business Failures in Addis Ketema Sub City:" A Case Study, Addis Ababa, Ethiopia, 2008.
- [19] Robin Burke, Hybrid Recommender Systems: Survey and Experiments California State University, Fullerton Department of Information Systems and Decision Sciences, 2002.
- [20] P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.), Case-Based Recommendation the Adaptive, pp. 342–376, 2007.
- [21] Priti Srinivas Sajja, Rajendra Akerkar, Knowledge-Based Systems for Development, 2010.
- [22] David Poole, Alan Mack worth, Randy Goebel, and Oxford University Press, Computational Intelligence, 1999.
- [23] J .Ranjan, P. Tripathi, Knowledge Acquisition, Representation and Reasoning, 2006.
- [24] E. Janet Burge, Knowledge Elicitation for Design Task Sequencing Knowledge, Worcester Poly Technic Institute, 1998.
- [25] R. lópez and de mántaras , Retrieval, Reuse, Revision, and Retention in Case Based Reasoning, 1998
- [26] Julie Main, Tharam Dillon and Simon ShiuA, Tutorial on Case-Based Reasoning, Soft Computing in Case Based Reasoning, pp.1-28, 2001.
- [27] Agnar Aamodt and Enric Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches, Vol. 7 Nr. 1, March, pp 39-59, 1994.
- [28] Christiane Gresse von Wangenheim, *Case-Based Reasoning A Short Introduction*, 2000.
- [29] Tagel Aboneh, "Knowledge Based System for Pre-Medical Triage treatment at Adama University Asella hospital," 2013.
- [30] Azeb Bekele Eshete, "Integrated Case Based and Rule Based Reasoning for Decision Support," Norwegian University of Science and Technology Department of Computer and Information Science, 2009.
- [31] Dereje Workie Nigussie, "Role of Financial Institutions in the Growth of Small and Medium Enterprises in Addis Ababa," 2012.
- [32] Abiyu Jiru, "Factors Constraining the Growth and Survival of Micro and Small Enterprises in Burayu," Addis Ababa, Ethiopia, 2011.
- [33] Minilek Kefale and K. P. M Chinnan, Employment Growth and Challenges in Small and Micro

- Enterprises Woldiya, North East Amhara region, Ethiopia Bahir Dar University, Ethiopia, 2012.
- [34] Habtamu Tezera, a Research Project on Challenges and Opportunities of Small and Medium Scale Enterprises (SMEs) in Addis Ababa (The Case of Arada Sub City), 2000.
- [35] Mohammad Darzi, Mahdi Hosseini, Ali Asghar Liaei, Zahra M. Manesh, Habibollah Asghari Accelerating Growth in High Tech MSEs through Tacit Knowledge Sharing Using Case-Based Recommender Systems, 1998.
- [36] Ralph Bergmann, Introduction to Case-Based Reasoning, uni-kl.de, University of Kaiserslautern, 2000.
- [37] Jim Prentzas and Ioannis Hatzilygeroudis, Categorizing Approaches Combining Rule Based and Case-Based Reasoning. Department of Computer Engineering and Informatics, School of Engineering, University of Patras, Greece, 2007.
- [38] Survey on Micro and Small Enterprises (MSEs) in Selected Major Cities of Ethiopia, Addis Ababa, June, 2013.
- [39] Admasu Abera, "Factors Affecting the Performance of Micro and Small Enterprises in Arada And Lideta Sub-Cities," Addis Ababa, October 2011.
- [40] Halil Tunal and Ferda Yerdelen Tatoğlu, "The analysis of factors affecting investment choices of Households in Turkey with Multinomial Logit model", *department of economics, faculty of Economics, Istanbul University, turkey, 2010.*
- [41] Justin Beneke, "Factors Affecting International Investment," Istanbul, Turkish, 2012.
- [42] Fauzilah Salleh and Mohamed Dahlan Ibrahim, "Demographic Characteristics Differences of Risk Taking propensity among Micro and Small Business Owners in Malaysia," Faculty of business Management and accountancy gong Badak Campus, Sultan Zainal Abidin University, 2011.
- [43] Olabisi, Sherifat Yusuff, "Factors Affecting Small-Scale Business Performance in Informal Economy in Lagos State-Nigeria": a Gendered Based Analysis Department of Sociology Lagos State University, 2009.
- [44] Ashok Kumar Panigrahi, Risk management in Micro, Small and Medium Enterprises (MSMEs) In India: a Critical Appraisal vol.1 (4), 2012.
- [45] Sems Ozar, Micro and Small Enterprises in Turkey: Uneasy Development, Bogazici University,

1998.

- [46] Wyattville Road and Loughlin stown, "Features of Micro and Small Enterprises in the EU Employment Relations in Micro and Small firms", 2001.
- [47] UNCTAD, *an Investment Guide to Ethiopia, Opportunity and Conditions*, Ethiopia, Addis Ababa, 2000.
- [48] *Micro and Small Enterprises Development Strategy*, Addis Ababa, Ethiopia, 2005.
- [49] E. De kock, Acquired Knowledge is organized so that it will be ready for use, in an Activity Called Knowledge Representation, University of Pretoria Etd, 2003.
- [50] A. Salem, "A Case Base Experts System for Diagnosis of Heart Disease", *International Journal On Artificial Intelligence and Machine Learning*, 5(1), pp. 33-39, 2005.
- [51] W. Clancey, " Knowledge Acquisition for Classification Expert Systems". *Proceedings of ACM National Conference on computer science*, Stanford University, 1984.
- [52] Nasrullah Iqbal and Muhammad Hassan Ashraf, "Evaluation of jCOLIBRI", 2006.
- [53] A. Juan, Recio-Garca and Belen Daz-Agudo, *An Introductory User Guide to jCOLIBRI*, Dep. Sistemas Informaticosy Programacion, Universidad Complutense de Madrid, Spain, November, 2004.
- [54] David Mc Sherry, Precision and Recall in Interactive Case-Based Reasoning: In Case Based Reasoning Research and Development (ICCBR), *Lecture notes in Artificial Intelligence*, pp. 392-306, 2001.
- [55] Henok Bekele, "A Case-Based Reasoning Knowledge Based System for Hypertension Management," Addis Ababa University, Ethiopia, 2011.

## APPENDIX

### **Interview questions to domain experts**

This is an interview which is prepared for the objective of gathering information for the selection of enterprise sectors and activities from the federal micro and small enterprise development agency enterprise experts:

1. What type of advising system is given when members are come to start an enterprises?
2. What are the criteria's used for assigning members to new enterprises?
3. From the criteria's/attributes which are more important to assign members in to enterprises?
4. Which enterprise sector is more suitable for females and males members?
5. Which enterprise sectors are used to get benefits by a long period of time?
6. Which enterprise sectors are used to give benefits by a short period of time?
7. How much capital is required to start an enterprise?
8. What is the influence of location for the success of enterprise sectors?
9. What are the main factors to the failure of enterprise sectors?

### **Interview questions for enterprise members**

After introducing the objective of the study and requesting the respondents' participation in the study the interviewer records their answers for the following interview questions.

1. In what way you get advising from the experts?
2. What are the problems in advising systems of FeMSEDA?
3. Is there any experienced enterprise expert who can give a brief advice for how to start an enterprise and where to start as well as how much amount of capital is needed?
4. How to select the enterprise sector and enterprise activity?
5. Have you get any guidance from experts to select enterprise sectors and enterprise activities?

### Prototype Evaluation form for the Domain Expert and members

This is an evaluation form to be filled by enterprise experts and members to evaluate the applicability of the case-based recommender system in enterprise sector and activity selection. I thank you in advance for your willingness and valuable time. Description of the parameter values are as follows.

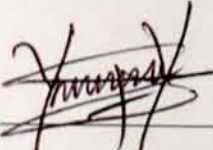
Performance value	1	2	3	4	5
Description	Poor	Fair	good	Very good	Excellent

**Instruction:** please assign (X) on the appropriate value for the corresponding parameter of evaluation questions of the case based recommender system in enterprise sector selection.

No	Evaluation criteria	Performance value					Average	Percentage
		1	2	3	4	5		
1	Is the system is efficient in time							
2	Easy to use							
3	Is the interface is interactive for users							
4	Relevancy of the retrieved cases for decision making							
5	Adequacy and clarity of decision support							
6	Relevancy of the attributes in representing members case							
7	Is the explanation facility is clear							
8	Significance of the system in the domain area							

## DECLARATION

This thesis is my original work and has not been submitted as a partial requirement for a degree of master in any other university.

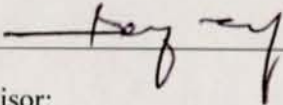


---

Yirga Yayeh

June, 2014

This thesis has been submitted for examination with my approval as a University advisor.



---

Advisor:

Dereje Teferi (PhD)