ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF INFORMATION SCIENCE

APPLICATION OF CASE BASED RECOMMENDER SYSTEM IN THE TOURISM SECTOR FOR THE SELECTION OF TOURIST ATTRACTION AREAS IN ETHIOPIA

BY: TAMIR ANTENEH ALEMU

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

June, 2014

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF INFORMATION SCIENCE

APPLICATION OF CASE BASED RECOMMENDER SYSTEM IN THE TOURISM SECTOR FOR THE SELECTION OF TOURIST ATTRACTION AREAS IN ETHIOPIA

BY: TAMIR ANTENEH ALEMU

Name and Signature of Members of the Examining Board

_____	_____	_____
Chair person, Examining Board	Signature	Date
Dr. Rahel Bekele _____	_____	_____
Advisor	Signature	Date
Dr. Tibebe Beshah _____	_____	_____
Examiner 1	Signature	Date
Mr. Getachew Jemaneh _____	_____	_____
Examiner 2	Signature	Date

ACKNOWLEDGMENT

First and for most I would like to give a special gratitude to the Glory of God and His mother Saint Mary who provided me the courage to finish my program of study.

I am deeply indebted to my advisor Dr. Rahel Bekele for her dedication, encouragement, inspiring guidance, and valuable comments throughout my thesis work.

Next to that, I am also extremely grateful for my former instructor Dr. Million Meshesha for his positive encouragement before the work was started in the proposal development phase and for his constructive guidance after the work has been started. My special thanks go to the National Tour operation and Ministry of Culture and Tourism domain experts who devoted their precious time for the interviews and consultation sessions throughout the research and for providing me valuable facilities and resources.

I give my warmest gratitude to, Mr. Michael Alem (tourism promotional expert at NTO) for providing relevant data cases, and other necessary information and encouraging me by giving necessary material for my research work.

My thanks also goes to Bahir Dar University (BDU) for granting me a study leave with the necessary benefits, without which I could not have been able to join my Msc. study here in Addis Ababa University (AAU).

My greatest gratitude is extended to all my family members for their encouragement and supports throughout my study. Finally, I would also like to thank sincerely all my friends who helped me with their valuable support during the entire process of this thesis.

Table of Contents

ACKNOWLEDGMENT I
LIST OF FIGURES VI
LIST OF TABLES VII
LIST OF ACRONYMS VIII
ABSTRACT IX
CHAPTER ONE 1
1. Introduction 1
1.1. Ethiopian Tourism 3
1.2. Statement of the problem and its justification 5
1.3. Objective of the study 7
1.3.1. General objectives 7
1.3.2. Specific objectives 7
1.4. Scope and Limitation of the study 8
1.5. Methodology 8
1.5.1. Data source 8
1.5.2. Data collection method 9
1.5.3. Feature selection 10
1.5.4. Knowledge (Case) Representation 11
1.5.5. Development Tools 12
1.5.6. Testing / Evaluation 12
1.6. Significance of the Study 13
1.7. Organization of the Thesis 14
CHAPTER TWO 15
2. Tourism in Ethiopia 15
2.1. Background 15
2.2. Historical Explanation of Tourist flows and Tourism Receipt 16
2.2.1. Ethiopia as Tourism Destination 16
2.2.2. Ethiopian Tourism policy 18
2.2.3. Essential Requirements for Tourism 19

2.3. Sectors in Tourism 19
2.4. Categories of Tourist Attractions 21
2.4.1. Natural Attractions 21
2.4.2. Cultural Attractions 22
2.4.3. Historical Attractions 22
2.4.4. Archaeological Attractions 23
2.4.5. Religious Attractions 23
2.5. Hierarchical structure of Tourist attraction 24
2.6. The Relationship between attractions and destinations 25
2.7. Tourist attraction areas in Ethiopia in Different Regions 26
2.8. Types of Visitors 30
CHAPTER THREE 32
3. Literature Review 32
3.1. Introduction 32
3.2. Definition and Basic concept of Recommender system 33
3.3. Architecture of Recommender system 35
3.4. Recommendation techniques 36
3.5. Types of Recommender system 37
3.5.1. Knowledge based recommender system 38
3.5.1.1. Knowledge based system (case based reasoning system) 39
3.6. Knowledge Engineering 41
3.6.1. The Knowledge-Engineering Process 41
3.6.2. Knowledge Acquisition. 43
3.6.2.1. Difficulties in knowledge acquisition 46
3.6.3. Knowledge Representation 46
3.6.4. Knowledge validation 47
3.6.5. Inference Engine 47
3.6.6. Explanation Module 47
3.6.7. Techniques of knowledge based Reasoning 47
3.6.8. Case Based Reasoning 48
3.7. The CBR cycle 53

3.8. CBR Techniques 54
3.9. Benefits of CBR 54
3.10. Limitations of CBR 55
3.11. The comparison of Rule based & case based reasoning 56
3.12. CBR system performance evaluation methods 57
3.13. CBR development Tools 58
3.14. Related works on Recommender system 61
CHAPTER FOUR 66
4. EXPERIMENTAL DESIGN 66
4.1. Knowledge Elicitation 66
4.2. Knowledge structuring 67
4.3. Knowledge Acquisition from Domain Experts 67
4.4. Knowledge Acquisition from Tourists 69
4.5. Knowledge Acquisition from Relevant Documents 70
4.6. Selecting an Attribute using Data Mining tools 71
4.6.1. Data collection 71
4.6.2. Data preparation and cleaning 71
4.6.3. Selecting Best Attributes (Features) 73
4.7. Knowledge Modeling 75
4.8. Factors that affects the selection of tourist attraction area decision 77
4.8.1. Gender 77
4.8.2. Age 78
4.8.3. Types of Visitors 79
4.8.4. Travel purpose: 79
4.8.5. Length of stay: 79
4.8.6. Location of attraction areas 80
4.9. The case structure of Tourist attraction area selection 81
CHAPTER FIVE 84
5. PROTOTYPE DESIGN AND IMPLEMENTATION 84
5.1. Introduction 84
5.2. Architectural Design of CBR system for Tourist Attraction Area Selection 84

5.3. Case Based Reasoning System for Tourist Attraction Area Selection 87
5.4. Creating new CBR Application 89
5.4.1. Case Base Building 91
5.4.2. Case Representation 93
5.4.3. Description of Case Attributes 94
5.5. Managing Connectors 98
5.5.1. Managing Tasks and Methods 100
5.5.1.1. Managing Tasks 101
5.5.1.2. Managing Methods 105
5.5.2. Similarity of cases, Matching and Ranking 107
5.5.3. Deploy the Case Based Recommender system 109
5.5.4. Explanation Facility 110
CHAPTER SIX 111
6. TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE 111
6.1. Introduction 111
6.2. User Acceptance Testing 111
6.3. Case Similarity Testing 116
6.4. Testing the CBR Cycles and Evaluating the Performance of the System 119
6.4.1. Evaluation of the Retrieval Process 119
6.4.2. Evaluation of the Reuse process 122
6.4.3. Performance Comparison of the recommender system with Previous CBR Systems 123
CHAPTER SEVEN 126
7. CONCLUSION AND RECOMMENDATIONS 126
7.1. Conclusion 126
7.2. Recommendations 129
Reference 131
Appendixes 136
DECLARATION 138

LIST OF FIGURES

Figure 2.1 hierarchical structure of attractions (source: Pamela, 2009) 24

Figure 2.2 classification of visitors (source: MoCT bulletin, 2010)..... 31

Figure 3.1 Architecture of recommender system (source: Terveen and Hill, 2001). 35

Figure 3.2 Recommendation techniques and their knowledge sources (source: Burke, 2007)..... 37

Figure 3.3: Architecture of knowledge based system 40

Figure 3.5 basic idea of CBR approach (source: Wangenheim, 2000)..... 50

Figure 3.6 the major components of CBR system (adapted from: Mani, 2001)..... 52

Figure 3.7 The CBR Cycle (source: Aamodt and plaza, 1994) ……..……..53

Figure 3.8 software architecture of jCOLIBRI(source: Antanassov, A. & Antonov, L. 2012). 60

Figure 4.1 hierarchical structure of tourist attraction area selection 76

Figure 4.2 location of attraction area……..……..79

Figure 5.1 the architecture of CBR in tourist attraction area selection process..... 86

Figure 5.2: The Main Window of jCOLIBRI 88

Figure 5.3 creating new CBR applications..... 907

Figure 5.4 categories of jCOLIBRI extention……88

Figure 5.5 CBR Applications..... 91

Figure 5.6 Case structure defining and similarity..... 95

Figure 5.7 JCOLIBRI connector schema (source: García, et al, 2008)..... 99

Figure 5.8 managing connectors..... 100

Figure 5.9 configure the CBR application..... 102

Figure 5.10 Revision and Retain tasks……..102

Figure 5.11 Managing tasks and methods..... 106

Figure 5.12 Case similarity between case base and query..... 108

Figure 5.13 windows for case entry in to the case base 109

LIST OF TABLES

Table 4.1 the case structure for tourist attraction area selection. 82

Table 5.1 description and weight of the selected attributes 96

Table 6.1 performance evaluation by domain expert 113

Table 6.2 performance evaluation by visitors 115

Table 6.3 Sample of queries that are used in this experiment with their values 117

Table 6.4: query similarity with their corresponding cases from the case base 118

Table 6.5: Relevant Cases Assigned by the Domain Expert for Sample Test Cases 120

Table 6.6 Depicts Performance Measurement of the system using Precision and Recall 121

Table 6.7: Accuracy value of the reuse process 123

Table 6.8: a comparison of the system with the previous CBR systems 124

LIST OF ACRONYMS

AI……Artificial Intelligence

CBR ……Case Based Reasoning

DB……Database

ES……Expert System

GAIA ……Group of Artificial Intelligence Applications

GDP……Gross Domestic Product

GUI……..Graphical User Interface

JCOLIBRI……java Cases and Ontology Libraries Integration for Building Reasoning Infrastructure

KA……Knowledge Acquisition

KB……Knowledge Base

KBS……..Knowledge Based System

KE……Knowledge Engineering

MoCT……..Ministry of Culture and Tourism

NTO……National Tour Operation

SQL……Structured Query Language

WTO……World Tourism Organization

Words Used interchangeably in This Paper

--Visitors and Tourists

ABSTRACT

The aim of this research is to design a prototype case based recommender system for tourist attraction area and visiting time selection that can assist experts and tourists to make timely decisions. For the development of case based recommender system, essential knowledge was acquired through semi-structured interview and document analysis. Eight domain experts and fourteen visitors were interviewed to elicit the required knowledge about the selection process of attraction area. The acquired knowledge was modeled using hierarchical tree structure and it was represented using feature value case representation. At the end, jCOLIBRI programming tool was used to implement the system.

The main data source (case base) used to develop case based recommender system for tourist attraction area selection is previous tourist cases collected from NTO and MoCT. As a retrieval algorithm, nearest neighbor retrieval algorithm is used to measure the similarity of new case (query) with cases in the case base. Accordingly, if there is a similarity between the new case and the existing case, the system assigns the solution (recommended attraction area and visiting time) of previous case as a solution to new case.

To decide the applicability of the prototype system in the domain area, the system has been evaluated by involving domain experts and visitors through visual interaction using the criteria of easiness to use, time efficiency, applicability in the domain area and providing correct recommendation. Based on prototype user acceptance testing, the average performance of the system is 80% and 82% by domain experts and visitors respectively. The performance of the system is also measured using the standard measure of relevance (IR system) recall, precision and accuracy measures, where the system registers 83% recall, 61% precision and 85.4% accuracy. Finally, conclusion and future research directions are forwarded.

CHAPTER ONE

1. Introduction

Nowadays it is very important for people to be supported in their decisions, due to the exponential increase of the available information. This exponential growth of information creates information overload. However, the term is defined, there cannot be many people who have not experienced the feeling of having too much information which uses up too much of their time, causing them to feel stressed which, in turn, affects their decision-making. i.e people may tend to be reluctant in making decision or they may leads to wrong decision (Angela and Anne, 2000). Recommender systems have proven to be an important response to such a problem by providing users with more proactive and personalized information services. It usually track user's behavior and collect information about items that the user seems to be interested in so that they can build a model of what user likes (Gemmis et al., 2009).

Recommender systems attempt to reduce information overload and retain customers by selecting a subset of items from a universal set based on user preferences. RS can be defined as "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" (Gemmis et al., 2009). It can also be described as an information filtering technology that is used to present information on items to the user that are in line with the users tastes. These systems involve predictive models, heuristic search, data collection, user interaction and model maintenance (Marla, et al., 2011).

A recommender system is also a tool that supports users in identifying interesting items especially among large numbers of items. They are used to support users in their decision-making in daily life situations in terms of pre-selecting information that might be of interest

to them, where they confronted situations without sufficient experience in the available alternatives (Ghauth and Abdullah, 2011).

Recommendation process is a sequential process. It is not 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 (Guy et al., 2010).

Recommender systems may be based on collaborative filtering (by user ratings), content-based filtering (by keywords), 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) (Burke, 2006, Schafer et al, 2001). Case based recommender system is a part of knowledge based recommender system that exploits case based reasoning to generate personalized recommendations for exploiting the knowledge contained in past recommendation cases. These systems assume that the quality of a new recommendation depends on the quality of the recorded recommendation cases (Burke, 2007). A case based recommender system maintains a set of cases of previously solved recommendation problems and their solutions. According to (Shimazu, 2002), case base is the product catalogue which is the solutions of the recommendation problem, and 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.

1.1. Ethiopian Tourism

Tourism- is the activities of persons traveling to and staying in a place outside their usual environment for not more than consecutive year for leisure, business and other purposes (Ministry of urban development and construction, 2009).

It is one of the largest and rapidly growing industries in the world, and is even considered by the UN World tourism Organization as the biggest industry in the world when related and complementary industries are taken into consideration. Obviously Tourism constitutes the back bone of the country economy. The Ethiopian Tourism sector plays a vital role in the industrial development of the country. Tourism makes a valuable contribution to the economic, social and environmental well-being of Ethiopians across the country. It is unique in its economic, environmental and socio cultural significance to every region while enabling Ethiopians to explore their heritage and celebrate their culture as they travel within the country (Yechale, 2011).

Ethiopia has immense tourism potential owing to its natural, historical and cultural endowments and the flow of tourists in the country becomes increasing from time to time. Tourism in Ethiopia dates back to the pre-Axumite period when the first illustrated travel guides to Ethiopia can be found in the friezes of the pyramids and ancient sites of Egypt. These depicted travels to the land of Punt, which the Egyptians knew was the source of the Nile, and where they traded for gold, incense, ivory and slaves. The fourth century Persian historian Mani described the Kingdom of Axum as being one of the four great empires of the world, ranking it alongside China, Persia and Rome (World Bank, 2006).

Tourism has emerged as one of the world's socio-economic sectors and has been steadily expanding at average rates of about 4 and 4.5 percent annually during the latter half of 20th century. Globally tourism generated an estimated US$3.4 trillion in gross output, contributing 10.9 percent of the world's gross domestic product (GDP), creating employment opportunity

for about 212 million people and producing US $637bellion in government tax revenues by the year 1995, (World Tourism Organization, 1995). Tourism has become one of the major deriving forces of the world economy with 903 million tourists traveling every year globally (UN WTO, 2008). Until recently tourism become as a means of solving developmental bottle necks of developing nations.

The tourist in flow in the developing countries has both benefit and cost. But most of the tourism programmers in developing countries have been carried out without sufficient and careful attention of the various benefits and cost involved. In this regard tourism has played both positive and negative roles in the developing countries. The positive impacts can be providing employment opportunities, generating foreign exchange, development infrastructures and social services, contribution towards the preservation of cultural heritages and developing cross cultural exchange. On the other hand, creating unbalanced economic development, the feeling of dependence on tourists, increasing incidences of crimes, loss of historical resources, aggravated prostitution, alcoholism, in sanitary condition and influencing the customs, lifestyle and tradition, environmental pollution and political influences of the host communities are among the few examples of the negative impact of tourism, (Ianranges, 2010). As Ethiopia is among these developing countries the above mentioned impacts directly or indirectly influence especially the main tourist destination of the country such as Axum, Gander, Lalibela, Bahir Dar and other tourist destinations areas of Ethiopia.

The Ethiopian Ministry of Culture and Tourism is the part of the Government of Ethiopia responsible for developing and promoting the tourist products of Ethiopia both inside the country and internationally. In doing so the Ministry closely works together with different national and international stakeholders. It publicizes the country's resources of tourist attractions and encourages the development of tourist facilities. It also licenses and supervises establishments of tourist facilities such as hotels and tour operators, and sets the standards for them (www.tourismethiopia.org).

1.2. Statement of the problem and its justification

There are a number of problems in Ethiopian Tourism faced by Experts in the sector and tourists alike. The problems faced by experts in the sector are lack of appropriate, relevant and understandable information that they need to give advice and guidance to their clients. As stated by (Culture and Tourism Office, 2011), advice is one of the most important problems of the tourism sector because the sector also uses traditional advisory system of tourists and the traditional nature of the existing advisory system make it interesting to undertake the study. Lack of access to appropriate information is encountered by Ministry of Culture and Tourism due to the fact that information system is not developed to enable proper collection, organization and dissemination in the sector.

In addition, there is no integration or collaboration among different experts that are found in different tourism sectors to develop an organized guidance to new tourists because collection of ideas from different experts is important to develop well defined and organized guidelines to the tourists. For instance, one expert may have awareness about natural tourist attraction areas but have no more idea about historical tourist attraction areas, etc. These shows experts advice is limited only with the one which is most familiar with them (United Nations World Tourism Commission, 2007).

According to (Ministry of Urban development and construction, 2009), the main problems in Ethiopian tourism sector advice system is, the advising services given are not fast. Because of this, it takes long time to get the required tourist attraction areas and permissions to visit. As to the knowledge of the researcher, no attempt has been made to avert this problem on the domain area. Difficulty of getting appropriate advice is a critical issue for tourists since knowing the right tourist attraction area is a key factor and also knowing the right time to visit in is another factor to consider to new tourists (Ministry of Culture and Tourism, 2012). As stated by Yechale (2011), Tourists tend to lose money by making the wrong choices of areas to visit and the wrong time without the consideration of their characteristics or location of the area or income level of the tourist.

In the context of visiting, the wise words of the oracle emphasize that success depends on ensuring that your visiting strategy fits your personal characteristics (Ministry of Culture and Tourism, 2012). Even though all visitors are trying to get satisfaction, each one comes from a diverse background and has different needs and capabilities. It follows that specific visiting vehicles and methods are suitable for certain types of visitors.

According to the interview made with Ministry of Culture and Tourism development promotion expert, tourists have a few factors to consider, when looking for the right place to visit such as age, nationality, gender, travel frequency, attraction preference and length of stay. The expert further comments that due to lack of knowledgeable domain experts to give appropriate advice in Ministry of Culture and Tourism, visitors are confused about where to visit, how much budget they have to allocate, and which tourist attraction area is best to them to be satisfied / successful in their recreational program.

In summary, providing an effective service in the Tourism sector of Ethiopia is critical to attract more foreign and local tourists. However, there are major problems that need immediate solution. First, the difficulty of getting fast, reliable, and consistent expert advice in the sector that is suitable to each visitor's characteristics and capabilities. Second, inadequacy of the number of experienced experts and consulting individuals who can give advice on tourism issues in the country. Therefore, this study aims to design a recommender system that helps visitors to get fast and consistent advisory service so that visitors can identify the best tourist attraction areas that have the highest potential of success / satisfaction and that match their personal characteristics.

To this end, this study attempts to answer the following research questions:

- What are the most important attributes expected from visitors to assign them in a specific tourist attraction areas?

- How does one acquire, model, represent, and implement the required knowledge for a recommender system?
- How can we develop an effective prototype of recommender system for tourist attraction area selection?
- What is the contribution of the recommender system over manual system for tourists in the selection of tourist attraction area?

1.3. Objective of the study

The following general and specific objectives are formulated towards solving the research problems.

1.3.1. General objectives

The general objective of the study is to design a prototype Recommender System that can provide possible recommendation on the selection of tourist attraction areas in Ethiopia to foreign and domestic visitors.

1.3.2. Specific objectives

- To identify the main criteria that influences the tourism sector in the selection of tourist attraction areas for new tourists.
- To develop a prototype of recommender system to visitors on the selection of tourist attraction areas that best matches with their characteristics.
- To determine the performance of the proposed recommender system using different evaluation techniques.
- To recommend further research areas for future work.

1.4. Scope and Limitation of the study

This study focus on developing a prototype recommender system for the tourism sector on tourist attraction area selection based on different characteristics of tourists. The recommender system does not include other activities of the tourism industry such as Conformity Assessment and Certification of Tourist Accommodation, Tourism fund, etc. Moreover, the aim of the research is to develop a recommender system that can provide suitable advice for foreign and local tourists in the tourism sector. Developing a full-fledged system demands one to construct maps of different levels for easy exploration, which in turn requires a long period of time.

The core limitation in this developed system is, the explanation facility is unable to provide response or feedback based on visitors' questions. It gives only a direct explanation when the system recommends a solution. For instance, if one visitor is not clear about length of time, there is no a possibility to ask explanation. The other limitation is, as previous tourist cases are collected from NTO, the case structure for the prototype is made up of using attributes that are recorded in the file of previous visitors. Due to these, attributes like: housing preference, purchasing habit, marital status, level of education, level of risk taker which is considered in tourist attraction area and suitable visiting time selection are not included in the case structure for the prototype.

1.5. Methodology

The following methodologies have been used in the course of this study to achieve the above stated objectives.

1.5.1. Data source

The main data source used for this study was domain Experts working at MoCT (Ministry of Culture and Tourism) and NTO (National Tour Operation) as well as previously solved cases which are available in the aforementioned organizations. The Researcher selects these

organizations since they use traditional advisory system of tourists and the traditional nature of the Existing advisory system make it interesting to undertake the study.

1.5.2. Data collection method

To collect the required domain knowledge, both primary and secondary data collection methods have been employed. As primary sources, Tourism experts from MoCT & NTO and tourists have been interviewed. In addition, relevant literature from all possible sources including journal articles, tourism related websites, manuals especially on Ethiopian tourism, and guidelines has been reviewed. To acquire the required tacit knowledge from the selected domain expert, the researcher has employed semi- structured interview technique which focuses on the concept, procedures, and guidelines as well experience which domain expert used while advising in tourists. The researcher selects semi-structured interview technique because it allows the interviewer to change the order of questions and add new question based on the participant response.

The primary source of dataset used to undertake this research was previous visitor's case from NTO office. The dataset contains a total of 21,347 visitor cases recorded in a year from 2006-2009. The dataset of each year was stored in a hardcopy form and the researcher tried to put some of them in to different Microsoft excel file. The researcher used purposive data selecting methods to select sample datasets for this research. Because there are large amount of data and encoding all of the data in to a computer is the toughest and time consuming task. Therefore, the main objective is to select successful visitors in each attraction areas and the researcher selected visitors who are successful or satisfied in their recreational program for this research as a case base. In the recorded dataset there is a remark column filled by experts marked as satisfied and not satisfied. Of these, the researcher selects successful visitors which are marked as satisfied in their recreational program for the development of recommender system. There were totally 1079 successful visitors from the total of 21,347 visitors. But since the data set having different problems such as missing value of majority attributes, redundancy, etc, the researcher selects six hundred fifteen (615) cases of visitors from total of

1079 successful visitors. Finally these 615 cases represented as case base in a CSV format that are used as previously solved cases.

1.5.3. Feature selection

In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many redundant or irrelevant features. Redundant features are those which provide no more information than the currently selected features, and irrelevant features provide no useful information in any context. Feature selection techniques are a subset of the more general field of feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are often used in domains where there are many features and comparatively few samples (or data points). The archetypal case is the use of feature selection in analysing DNA microarrays, where there are many thousands of features, and a few tens to hundreds of samples. Feature selection techniques provide three main benefits when constructing predictive models:

- ✓ Improved model interpretability,
- ✓ Shorter training times,
- ✓ Enhanced generalization by reducing over fitting.

Feature selection is also useful as part of the data analysis process, as it shows which features are important for prediction, and how these features are related. Therefore, for this research, previous visitors' case was collected from NTO which contain eighteen (18) attributes and 21,347 cases. Since all the attributes are not equally important, the researcher selects important features which is used for tourist attraction area and visiting time selection with the consultation of domain experts and using attribute selection algorithm in WEKA software. At the end, nine (9) attributes were identified and used for the prototype development of the proposed case based recommender system.

1.5.4. Knowledge (Case) Representation

After the knowledge is acquired, the next task is knowledge (case) representation. Although there are various knowledge representation methods, like relational database knowledge representation, feature-value case representation, predicate based representation and soft computing knowledge representation methods have their own advantages and disadvantages. But for this research the researcher used feature-value case representation. The reason for representing the cases using feature-value representation is that this approach supports nearest neighbor retrieval algorithm and it represents cases in an easy way (Salem, et al., 2005). This approach also uses old experiences to understand and solve new problems. It also reuses its solutions and lessons learned for future use. In addition, it represents cases in an easy way by using attribute and value pair representation (Bergmann, et al (2005), & Bullinaria A. (2005). The algorithms used to calculate the similarity of cases in a case base representation for this research are 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, it selects the most similar stored cases to the query.

In the process of case based recommender system development, Knowledge representation is one of the basic steps. It refers to the formalism, both syntax and semantics, used to store knowledge in the architecture. It is also the process of interpreting domain knowledge into computer understandable form using various knowledge representation techniques. The object of a knowledge representation is to express knowledge in a computer tractable form, so that it can be used to enable our AI agents to perform well (Bullinaria, 2005). The common Knowledge based reasoning techniques include semantic network, logics, rules, case base and frames. Among these, the researcher use case based representation method for this research because of its unique capabilities like: suitable and efficient in knowledge creation, easy for knowledge acquisition even the knowledge can be gained from recorded historical data and its self updatability nature over the other reasoning techniques (Aamodt, A., and Plaza, E. 1994).

1.5.5. Development Tools

To develop a recommender system there are various programming tools which are available both freely and commercially. Among this SWI-prolog, myCBR, and jCOLIBRI are the most widely used and known frameworks for teaching and academic research purpose. All of the aforementioned tools have their own capabilities and limitations.

According to (Juan, et al., 2009), jCOLIBRI framework has the following features. Therefore for this research the researcher used jCOLIBRI framework due to the following unique capabilities of the tool.

- jCOLIBRI supports the full CBR cycles (Retrieval, Reuse, Revise and Retain).
- jCOLIBRI is extensible, reusable, different types of users and different purposes (development, research and/or teaching), compatible with commercial applications and, supporting different types of CBR systems, since it is just a .jar file suitable for web applications.
- It is suitable for developing large scale applications.
- jCOLIBRI works well in external database.

1.5.6. Testing/ Evaluation

Once the prototype is developed, the functionality and user acceptance of the system should be tested. The evaluation processes focus on system's user acceptance of the prototype and the performance of the system. User acceptance measurements are concerned with issues how well the system addresses the needs of the user, whereas performance measurement determine if the system perform the required task successfully. In addition to this, the standard measures of relevance (performance of the system) in the information retrieval (precision, and recall) have been used to evaluate the performance of the prototype.

The researcher tested user acceptance of the system by involving twenty two (22) evaluators using visual interaction methods together with questionnaire. 8 evaluators were selected from domain experts purposely and 14 evaluators were from visitors. These system evaluators

were interacting with it by using appropriate cases. That is, forty one (41) sample cases has been selected purposely and then evaluators from the domain area were interacting with the system by taking a sample of test cases then, an experiment was conducted to know how new cases are matched with the cases from the case base using case similarity measurement. Each case are selected purposively and used to test the performance of the prototype. Based on that, they evaluate the performance of the system by using close ended questions. Recall and precision value of the system have been calculated based on its retrieval results.

1.6. Significance of the Study

From this study, primarily Ethiopian tourism office, tourism experts and tourists, specifically the inexperienced tourism expert professionals are the immediate beneficiaries to enhance their day to day activities. The prototype has a great significance to teach primary tourism experts and tourists in order to have well understanding about success and failure of tourism. Consequently, those tourism experts can use the system in recommending tourist attraction areas to tourists based on their personal or economical characteristics in the sector where highly qualified tourism expert professionals are unavailable. The developed prototype recommender system is used to provide advising services for both foreign and local tourists. The recommender system is developed using the knowledge of different domain experts and documentary sources which is used as organizational memory. Therefore, it gives better recommendation services where highly qualified tourism expert professionals are not found. Moreover, it is an academic exercise to fulfill the requirement of masters program that the researcher is enrolled in.

1.7. Organization of the Thesis

The study is organized into seven chapters. **Chapter one** is the introduction part, which contains background of the study, statement of the problem, objectives, scope and limitations of the study, significance of the study and methodology to carry out the research. **Chapter two** narrates on the development of tourism in Ethiopia. **Chapter three** discusses about review of related literature on the knowledge based recommender systems, about its background, architecture, development phases, and knowledge based system overview and its related issues are presented. **Chapter four** discusses the experimental design parts (the knowledge acquisition, representation and conceptual modeling procedures). **Chapter five** deals with implementation of case based recommender system. **Chapter six** presents the results found in the evaluation and testing process of the prototype case based recommender system. Finally, **chapter seven** focuses on conclusion and recommendations.

CHAPTER TWO

2. Tourism in Ethiopia

2.1. Background

Ethiopia, one of the countries on the African continent, is among many countries on the globe that have newly started to develop tourism practices (Jauhojärvi, 2011). Ethiopia's tourism potential is diversified: natural attractions that include some of the highest and lowest places in Africa along with immense wild life including some endemic ones; a very old and well preserved historical traditions with fascinating stelae, churches and castles to witness that, an attractive cultural diversity of about 80 nations and nationalities; and various ceremonies and rituals of the Ethiopian Orthodox Church which open a window on the authentic world of the old Testament(www.tourismethiopia.org).

Today's fast growth and spread of tourism may wrongly imply that the term is well known. However, tourism, as several scholars agree, is a sector of any economy which has not yet obtained one single definition of its own. This is turn may influence the way how to approach when studying the sector. Tourism, as it is the case around the world, is a wide spread and ubiquitous aspect of an economy: that it is rare that people from every corner of the world that do not recognize tourist every day in their vicinity. And yet tourism remains a term that is susceptible to diverse interpretation (Sharply, 2006). According to the same author tourism could be defined from two main groups or classification of definitions those are technical and conceptual definitions. Technical definition of tourism interprets tourism as the activity of a tourist defined as someone who travels for 24 hours or more outside his country of residence. In this definition the type of travelers identified are those who travel for business, for pleasure, health or other purpose. Also includes those who stay in destination area for less than 24 hours which are known as excursionist. Conceptual definition of tourism given from anthropological perspective sees tourism from the person that is perceived to be a tourist.

2.2. Historical Explanation of Tourist flows and Tourism Receipt

Ethiopia's great potential for tourism development is mentioned everywhere and no need of going to its details. It suffices to say that it has almost all types of primary tourist products: historical attractions, national parks with endemic wild life and cultural and religious festivals. UNESCO recognizes eight world heritage sites (as many as Morocco, South Africa and Tunisia and more than any other country in Africa): Axum's obelisks, the monolithic churches of Lalibela, Gondar's castles, the Omo Valley, Hadar (where the skeleton of Lucy was discovered), Tia's carved standing stones, the Semien National Park, and the walled city of Harar(Mulualem, 2010).

Tourism in Ethiopia dates back to the pre-Axumite period when the first illustrated travel guides to Ethiopia can be found in the friezes of the pyramids and ancient sites of Egypt. These depicted travels to the land of Punt, which the Egyptians knew was the source of the Nile, and where they traded for gold, incense, ivory and slaves. The fourth century Persian historian Mani described the Kingdom of Axum as being one of the four great empires of the world, ranking it alongside China, Persia and Rome (World Bank, 2006).

As stated by Mulualem (2010), Modern tourism in Ethiopia can be said to have started with the formation of a government body to develop and control it in 1961: the Ethiopian Tourist Organization. The earliest analysis on the tourist flows and expenditures in Ethiopia was done by UNESCO (1968).

2.2.1. Ethiopia as Tourism Destination

Ethiopia is one actor of the tourist destination region in Africa and she is a land of dramatic contrasts. Altitudes span from the lowest point of the African continent to the fourth-highest peak. Far from being the mountainous thirst land of Western myth, the Southern and western highlands of Ethiopia boast the most extensive indigenous rainforest to be found anywhere in the eastern half of Africa. The rift valley south of Addis Ababa, the capital city has a characteristically African appearance, with vegetation dominated by grass and flat-topped acacia trees In terms of mammalian abundance, Ethiopia is one of Africa´s key bird watching

destinations with a rapidly growing national checklist of more than 800 bird pieces including 16 endemics, as well as a similar number of near-endemics whose range extends into a small part of neighboring Eritrea and Somalia. Ethiopia´s fauna and flora, through essentially of sub-Saharan Africa display some strong links to lands north of the Sahara (Philip, 2009).

Attractions that have been and are visited over the years are mainly culture, history, nature, and wildlife (including bird watching). Recently, tracking has become popular and is one of the experiences and of great satisfaction for many visitors. "Much of Ethiopia´s fascination lies in its myriad historical sites and tourism revolves mainly around historical sites because tourism to Ethiopia revolves around historical sites, and Ethiopians identify strongly with their history and they generally enjoy speaking to visitors who share their enthusiasm (Philip, 2009). In the author`s opinion, the large number of ethnic groups with their own language, custom, tradition and culture has made the country's cultural attraction popular and because Ethiopia is the oldest independent nation in Africa and have a heritage dating back to the first century, Ethiopian historical route has become a unique attraction.

As all the other routs, visiting historic route can take short or longer time depending on the visitor`s needs to stay, spend and means of transportation selected. There are local airports and Ethiopian Airlines flies to all historical attractions daily. "A visitor can cover all the historic sites, Bahir Dar and drive to Blue Nile, Gondar to visit the 17th century castles of the Medieval Capital, the Rock-Hewn churches of Lalibela, Historical relics of the ancient capital just in five days"(Tourism Ethiopia, 2013).

Even though the number of tour operators in the country has grown in the past years, there are attractions that are undiscovered or discovered but not experienced. This ancient destination has many unseen attractions and is the only sub-Saharan country in Africa tangible historical remnants, because of poor performance and lack of promotion, the visited attractions have been less than half of what should have been visited. According to the world's tourism report, the visited Ethiopian tourist attraction does not cover more than 45 percent. This has happened because factors like transportation, hotels, cars, and well trained

tour guides were not available in the old days. The other factor that has buried Ethiopian tourism has been lack of promotion about the historical places found in the country and limited information posted on the websites, in brochures or fliers. (Lambadina tours, 2009).

The concept of Tourism was introduced to Ethiopia for the first time in 1961 By Mr. Habte Selase Tafesse, the man who is known as “a man who invented tourism in Ethiopia” (Ethio-American Trade and Investment Counsel 2011). Even though the tourism sector was recognized as a sector for economic growth in 1965, the country`s social, political and economical development was not stable and the share of the tourist flow was at less than one present in the year 2007(Mulualem, 2010). Ethiopia was registered as a member state of World Tourism Organization UNWTO in 1975, and it has the share of tourist flow to the East African region of seven countries (Tanzania, Kenya, Uganda, Djibouti, Eritrea, Somalia and Ethiopia (World Tourism Organization, 2009).

The Ministry of Culture and Tourism was established in 2005 by The Government of the Federal Democratic Republic of Ethiopia who has recognized the necessity of creating a strong government organ to lead the sector (Tourism development policy 2009). There is also fast development of the tourism sector took place and the number of international arrivals has increased starting from the year 2007.

2.2.2. Ethiopian Tourism policy

In order to work on reversing the situation, the Federal Democratic of Ethiopia has launched its first tourism policy with a vision “ to see Ethiopia`s tourism development led responsibility and sustainability in order to contribute its share to the development of the country by aligning itself with poverty elimination”.(Tourism development policy 2009). The tourism development policy made it clear that it has not been possible for the country to drive full benefits from the sector, and development has remained uncoordinated and unsustainable because of the absence of a clear policy that would lay the direction for the cooperation and coordination that should exist among all stake holders. Therefore, the policy was expected to be the main motivator for the movement and it is believed to be used by the

government as well as all stakeholders to revive and use the full benefit of the sector (Tourism development policy, 2009).

2.2.3. Essential Requirements for Tourism

According to the MoCT Bulletin (2010), critical requirements for Tourism were stated as here under:

A. **Time**, as the hours for leisure increase so does the opportunity for travel. Changes in work days or hours, school calendars will affect how and when people can travel. The overall travel pattern has moved from a two week vacation to 6-8 three or four day mini-vacations per year.
B. **Money**, the majority of travel requires ***discretionary income***. Discretionary income is money left over after all monetary obligations (food, rent and taxes) have been paid.
C. **Mobility** is the access to transportation (car, bus, plane, train or ship) and the hours required to get to their destination.
D. **Motivation** is the reason people travel. Motivations may include seeking novelty, education, meet new people, adventure or stress reduction.

2.3. Sectors in Tourism

Tourism is a collection of activities, services and industries that delivers a travel experience, including transportation, accommodations, eating and drinking establishments, retail shops, entertainment businesses, activity facilities and other hospitality services provided for individuals or groups traveling away from home. The World Tourism Organization (WTO) claims that tourism is currently the world's largest industry with annual revenues of over $3 trillion dollars. Tourism provides over six million jobs in the country, making it the country's largest employer. Another scholars Mathieson and Wall (1982) define it as "a temporary movement of people to destinations outside their normal places of work and residence, the activities undertaken during their stay in those destinations, and the facilities created to cater to their needs." According to Macintosh and Goeldner (1986) tourism is also "the sum of the

phenomena and relationships arising from the interaction of tourists, business suppliers, host governments and host communities in the process of attracting and hosting these tourists and other visitors."

The tourism industry in Ethiopia has been divided into eight different sectors or areas. The following sector descriptions are brief overviews. Accommodation, Adventure Tourism and Recreation, Attractions, Events and Conferences, Food and Beverage, Tourism Services, Transportation, and travel Trade (Pamela, 2009).

- ✓ **Accommodation:** Any provisions for overnight stays and related service usually housing on a per diem / week basis.
- ✓ **Food and Beverage**: Any provisions for meal or beverage service and the related manufacturing or service industry.
- ✓ **Transportation:** Any service which provides for the movement of people within or between tourist markets and the related manufacturing or service industry.
- ✓ **Events and Conferences:** Any organized program with a start and end date, designed to bring visitors together for a common purpose... this may be business or entertainment-related and include all directly-related planning, and management functions.
- ✓ **Attractions:** Any natural or manufactured facility which draws on the interest of visitors... not always revenue producing but may have maintenance costs. Tourist attractions can also be defined as "Natural sites, man-made facilities, businesses or destinations of provincial scope / interest that generate visitation from outside the immediate / local area; by offering outdoor, educational, scientific, natural, cultural, heritage or entertainment experiences. Its primary purpose is to provide visitors with an experiential product designed to satisfy the traveling needs of visitors but where the sale of goods is of a secondary nature."
- ✓ **Adventure Tourism and Recreation**: Any provisions for the active involvement of visitors in a tourism market.
- ✓ **Travel trade**: Any part of the industry involved in the direct sale or marketing of visitor services in any of the other industry sectors.

- ✓ **Tourism Services:** Provide support services for the tourism industry… not usually 'front-line' and rarely have contact with visitors

2.4. Categories of Tourist Attractions

As stated by Ministry of Culture and Tourism (2012), the categories of attractions, with most promising key opportunities for potential visitors exist in the country today are grouped in such a way:

2.4.1. Natural Attractions

It is an attraction that has been created by nature. Many of these areas have been given a status to protect their environment and provide facilities so that the public are able to enjoy the sights. There are attractions such as caves, waterfalls, seashores and any other scenic view interest that haven't been created by mankind. Ethiopia has been very forward-looking in its provision of national park areas and there are at present a dozen regions within the country that have been designated as protected areas for wildlife.

- ✓ **Semien mountain national parks**: This is home to the endemic mammals of Walia Ibex, Semien Fox, Gelada Baboons, Nyala and many species of birds and plants apart from its spectacular scenic beauty.
- ✓ **Omo National Park:** This is the largest in the country and endowed with Buffalo, Elephants, Giraffe, Cheetah, Lion, Leopard, Burchell's Zebra.
- ✓ **Abijatta-Shalla Lakes National Park:** it is situated in the Great Rift Valley, only 200 kilometers (124 miles) south of Addis Ababa which has two different lakes in one park i.e Abijatta and Shalla.
- ✓ **Awash National Park:** is one of the finest reserves in Ethiopia Lying in the lowlands east of Addis Ababa, and striding the Awash River. The dramatic Awash Falls as the river tumbles into its gorge is the site not to be missed in the national park. A special attraction is the beautiful clear pools of the hot springs (Filwoha). Forty-six species of animals have been

identified here, including Beisa Oryx and Swayne's Hartebeest. The bird life is prolific especially along the river and in amongst the 392 species recorded.

- ✓ **Dallol Depression:** Dallol is in the Danakil depression is the lowest point and the hottest place on earth. It is also the world's only below sea level land volcano.
- ✓ **Sof Omar:** Sof Omar is Situated 510km south of Addis near the town of Arba Minch, in between Lakes Abaya and Chamo. Animals to be seen are Bushbuck, Swayne's Hartebeest, Burchell's Zebra, Grant's Gazelle, Guenther's Dik-dik, Greater Kudu, Crocodile, Anubis Baboon, Grey Duiker. Birds seen include Red-billed Hornbill, Grey Hornbil,l Fish Eagle, Kori Bustard, Abyssinian Ground Hornbill..
- ✓ **Mago National Park:** Mago National Park is 2,162 sq km, 770km southwest of Addis Ababa, on east bank of Omo River. The highest point is Mount Mago. Mainly grass savannah, some forested areas around rivers. 56 species of mammals were identified there: buffalo, giraffe, elephant, lelwel hartebeest, lion, cheetah, leopard, zebra, gerenuk, oryx.

2.4.2. Cultural Attractions

Ethiopian culture is very multi-faced, reflecting the ethnic diversity of the country. In Ethiopia, the culture of almost all regions like Tigray, Amhara, Konso, Somali, Oromo, Harar, Afar, Gambela, SNNP, are quite different one from the other and fascinating one to visit. There are also various cultural attractions of Ethiopian festivals that must be visited by visitors. The most prominent one are: “Enkutatash” Ethiopian New Year, **“**Timkat” the feast of Epiphany, “Meskel” the finding of true cross, Feast of Saint Gabriel (Kulubi), and Fasika (Easter).

2.4.3. Historical Attractions

Ethiopia is a landlocked country in Eastern Africa located on the Horn of Africa. It is bordered on the north and northeast by Eritrea, on the east by Djibouti and Somalia, on the south by Kenya, and on the west and southwest by Sudan. As Africa's oldest independent country, Ethiopia covers 1.13 million sq km (437,794 sq miles). The country has various historical sites which have a great potential of attracting visitors to it. The top historical

tourist destinations of the country are: Bahir Dar (Blue Nile Falls and Lake Tana including ancient monasteries), Lalibela (The 11 rock-hewn churches in the 12th century, one of the world's most incredible man-made creations, they are a lasting monument to man's faith in God. Most travel writers describe these churches as the "eighth wonder of the world". These remarkable edifices were carved out of a solid rock), Gondar (Castle of Fasiledes is an impressive one in the town of Gondar), Axum (Tall granite obelisks of Axum which announces the adoption of Christianity by a 4th-century, Axum, Yeha Temple and Arc of the Covenant Church in Axum), Harar (The old City Wall is the main attraction and symbol of Islamic architecture and the hyena man who feeds hyenas on the outskirts of the town every night is another attraction), and Mekele(Emperor Yohannes palace and the monastery of Debre damo is the most prominent attraction areas).

2.4.4. Archaeological Attractions:

Lucy, 3.5 million years old, and the recent discovery Ramides, 4.4 million years old hominid fossil, are discovered in Haddar, along the Awash River, east of the country. They completed the missing link between Apes and men. **Hominids:** New fossils discovered in the Afar desert of eastern Ethiopia are a missing link between our ape-man ancestors some 3.5 million years ago and more primitive hominids a million years older, according to an international team led by the University of California, Berkeley, and Los Alamos National Laboratory in New Mexico. The fossils are from the most primitive species of Australopithecus, known as Au. Anamnesis, and date from about 4.1 million years ago. Awash and Omo valley are also part of the archaeological attractions.

2.4.5. Religious Attractions

Ethiopia is also endowed with various religious attractions in different parts of the country. The well known ones are: Axum Tsion, the rock hewan churches of lalibela, monastery of Debre damo, various monasteries that are found in islands of Lake Tana etc.

2.5. Hierarchical structure of Tourist attraction

As described by Pamela (2009), Attractions are the most important elements of a tourist destination as they provide the main reason or motivation for tourists to visit a destination. There is a large variety of tourist attractions and some of these are shown in the following chart:

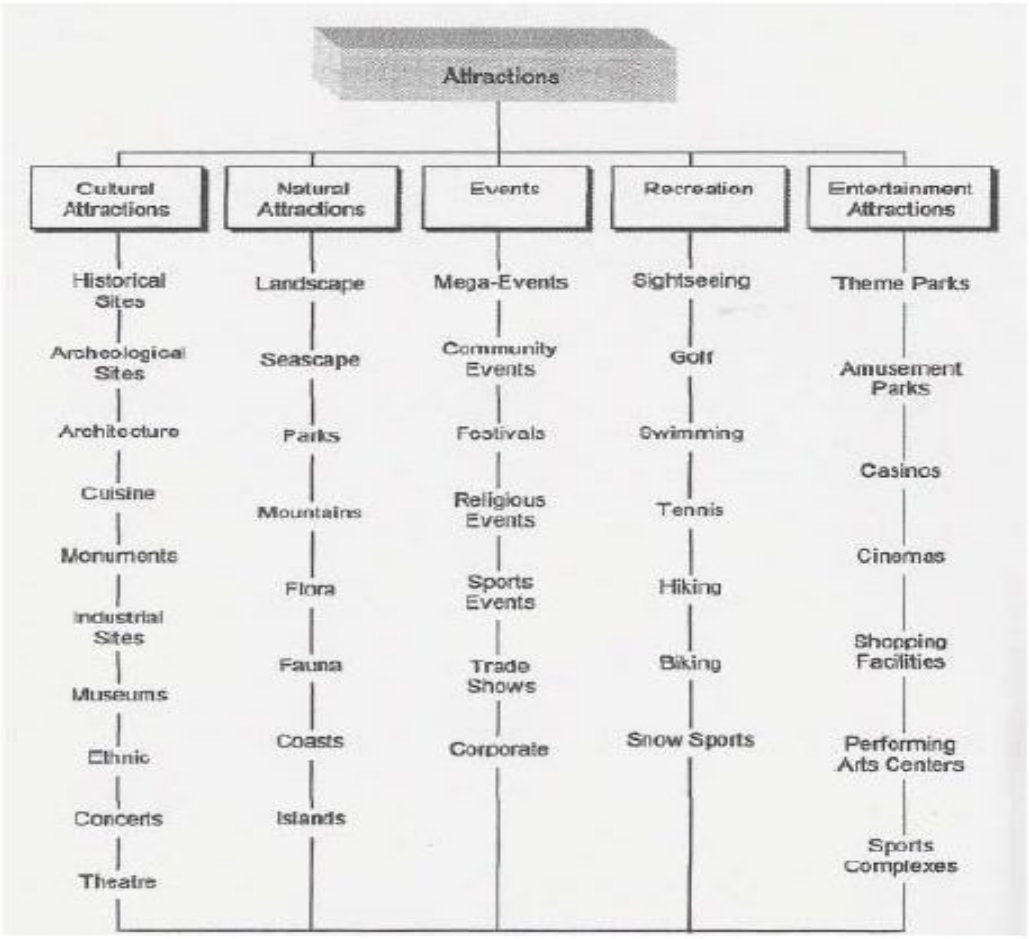


Figure 2.1 hierarchical structure of attractions (source: Pamela, 2009)

In general, a visitor attraction tends to be an individual site in a clearly defined area that is publicly accessible. The attraction motivates large numbers of people to visit it, usually for leisure, for a short, limited period of time. Any feature of a destination which attracts visitors, including places, venues or activities, can be called an attraction. Attractions usually have the following characteristics (Pamela, 2009):

a. Set out to attract visitors, including locals and tourists, who are managed accordingly?
b. Provides pleasurable and enjoyable experiences for visitors to spend their leisure time.
c. Developed to make it attractive and inviting for the use and enjoyment of visitors.
d. Managed as an attraction to satisfy visitors.
e. Provides facilities and services to meet and cater to the needs of visitors.
f. May or may not charge a fee for admission.

2.6. The Relationship between attractions and destinations

Attractions are generally single unit, individual sites with easily defined geographical areas based on a single key feature whereas destinations are usually larger areas that include many attractions with support services and infrastructure such as transportation networks and accommodation. There is a strong link between attractions and destinations. On one hand, a major attraction makes a destination more appealing to tourists and can stimulate the development of other tourism sectors such as hotels, tour operators and catering, as well as the destination itself. On the other hand, a popular and well-known destination ensures the potential market for the attractions. Destinations with high accessibility and clear market image are usually good locations to develop and build an attraction (Pamela, 2009).

As described by Pamela, (2009), Tourists are more likely to visit destinations that possess a wide variety of interesting facilities and services which they can enjoy. We can often find different kinds of attractions in a destination providing visitors with different types of experience. Some of these attractions are natural while the others are man-made. They can be broadly divided into four main types:

- ✓ ***Natural features:*** Physical features and natural scenery, collectively termed "landscapes", are major attractions for tourists who love nature. With the growing concern about conservation, environmental protection, landforms, natural vegetation and wildlife, natural features provide valuable resources for the development of nature-based travel and / or eco-tourism.

- ✓ ***Man-made buildings, structures and sites that were originally designed for a purpose other than attracting visitors:*** Attractions that were built to serve purposes other than attracting visitors may either be deliberately converted into an attraction or have spontaneously evolved into an attraction over time.
- ✓ ***Man-made buildings, structures and sites that are purpose-built to attract visitors and cater for their needs:*** The aim of purpose-built attractions is to attract visitors and increase visitor numbers. Satisfying visitors' needs is essential in the daily operations of these attractions.
- ✓ ***Special events:*** Festivals and events are one of the fastest-growing segments in tourism. Events are temporary attractions which provide opportunity for leisure, social or cultural experiences outside the normal range of daily activities.

2.7. Tourist attraction areas in Ethiopia in Different Regions

Since 1995, Ethiopia is divided into nine ethnically-based regional states and two chartered cities. These administrative regions replaced the older system of provinces. The word "kilil" more specifically means "reservation" or "protected area" (Wikipedia).

According to the manual of Ministry of Culture & tourism (2012), In Ethiopia there are nine regions (Location of attractions) and the 9 regional states or *kililoch* are based on ethnic territoriality. The regions with their respective attractions are listed as follows: Afar, Amhara, Benishangul-Gumuz, Gambela, Harari, Oromia, Somali, Southern Nations, Nationalities and peoples' region, Tigray as well as additionally there are two chartered cities such as Addis Ababa Dire Dawa.

A. Addis Ababa

Addis Ababa is a city administration and has 10 sub cities. As well it is the capital city of Ethiopia. There are various tourist sites that are found in Addis, such as museum of Galleries (Ethnological museum, 'Red terror' martyrs memorial museum and national museum of Ethiopia), Religious sites like st. George cathedral & museum, Adadi Mariam, Holy trinity

cathedral, Markets & Bazars like Merkato and finally Mount Entoto & Lion Zoo are the common attraction areas in Addis Ababa.

B. Dire Dawa

Dire Dawa is the second city administration in Ethiopia. Although Dire Dawa is primarily known by its trading centers, it also has its own numerous and precious cultures and heritages of glorious attractions. These places of interest are described hereunder: Porc Epic cave, Laga-oda ancient cave, Hinkuftu cave, Goda-Ajewa Cave, Abeyazid Mosque, Stale, Ancient Catholic Church, Italian Mosque, the Railway Station, and Italian Fort (Mishig).

C. The Oromia Regional state

It is one of the nine ethnically-based regional states of Ethiopia which has 18 zones. The Region is known as a land of Great Ecological Diversity. Currently several attractions are available in the region. Among these: Sof Omar Cave, Dire Sheik Hussien Mosque and Bale mountain park are registered as global heritage by UNESCO, and the largest portion of Ethiopia's rift valley Lakes Chain /Ziway, Langano, Abijata, Shalla, the highest pick mountain 'Mount Batu', the oldest park from the Continent Africa, Menagesha-Suba State Forest Park are also found there in the region. Oromia is the only region that sells three species for Trophy hunting from the world such as: The Bleeding Heart Baboon, Mountain Nyala, Menelik's Bushbuck.

D. **The Regional State of Amhara**

The Amhara National Regional State is the third largest state in the country. The region is blessed more with abundance and diversify natural environment than other regions in Ethiopia. It has more of the tourist attraction locations within its boundary. The major tourist attraction locations have been grouped into four according to Bureau of culture and tourism in Amhara region (Tourism Commision, 2005) as follows: Simen Mountain National Park, Rock-Hewn Churches of Lalibela listed as one of the World's heritage by UNESCO, Fassil-Gebbi Castles of Gondar is another World's heritage site, and Bahir Dar, the capital city of Amhara regional state serve as the gateway to the main attraction location in the region with

several tourist sites which include: Lake Tana , Islands and Peninsular Monasteries, Blue Nile falls, Wanzaye Hot spring, Orthodox churches , cultural activities, Souvenirs and artifacts of silver, brass and gold.

E. The Tigray Regional State

The State of Tigray consists of 4 administrative zones, one special zone, and 35 woredas. In the region, attraction areas are clustered in to five sites like Axum, Wukro, Cheralta, Mekelle and Maichew. These four cluster sites contain various cultural, religious and historical attractions, for example in the Axum cluster, as a religious heritage, Saint Yared: The Father of the liturgy in the Ethiopian Orthodox Church, Axum Tsion, Arabtu Ensessa Church and Gobo Dura (one of the great mount) is found. Tigray, Ethiopia's northern most regions has more than 120 rock-hewn churches found in a scattered manner unevenly; such as: Debretsion (Abune Abraham), Mariam Korkor, Yohannes Maequddi, Abune Yemata (Guh), Abune Gebre michael, etc.

F. The Regional State SNNPR

The State of Southern Nations, Nationalities and Peoples, is administratively divided in to 9 zones, 72 woredas and 5 special woredas. There are a total of 77 woredas and 149 towns in the State. Some of the major tourist attractions of the State are lakes like Awassa, Abaya and Chamo. Tropical forests such as Kaffecho, Shekecho and Omo zones are wonderful places. The Nechsar, Mago and Omo national parks and Konso culture center are found in this state. These national parks are centers of tourist attractions.

G. The Regional State of Somalia

The State of Somalia has a very large area size ranking second next to Oromiya. At present the state comprises 9 administrative zones and 49 woredas. Somalia has a number of local attractions, consisting of historical sites, beaches, waterfalls, mountain ranges and national parks. The famous National parks that are found in the region are the following: Daallo

Mountain, Hargeisa National park, Hobyo grasslands and shrublands, Jilib National park, Kismayo National park, Lag Badana National park.

H. **The Regional State of Harari**

Harari is one of the most popular historical towns in the Eastern part of Ethiopia. The State has no administrative zones or woredas and the total numbers of kebeles in the city are 19. The major tourist attractions in the region are stated as follows: The walled city of Harar with its five gates, Cultural houses, Alleys ways (streets), Historical sites, Shrines and mosques, Colorful women dress, Hyena feeding, Cultural museums, Crafts/basketry etc.

I. **The Regional State of Gambella**

The State of Gambella is composed of two administrative zones and eight woredas. Gambella national park is safari in the western savanna lowlands of Ethiopia. The fauna of the region include Elephants, Buffaloes, Monkeys and Parrots. Hot spring and mineral waters, waterfalls, densely natural forests are among the few resources and tourist attractions of the region.

J. **The Regional State of Afar**

The Afar Regional State (Afar: Qafar; Tigrinya: ዓፋር; Amharic: አፋር?) is one of the nine regional states (kililoch) of Ethiopia, and is the homeland of the Afar people. The State of Afar consists of 5 administrative zones, 29 woredas and 28 towns. Awash natural reserve, Yagundu-Ras national park and the Dallol depression are expressions of Ethiopia's desert beauty. Some of the attractions of this game reserve include Abyssinian wild ass, Grevy's zebra, beisa oryx, crocodiles, lions, grater kudu, wild (bat eared) fox, wild cat, cheetah, Grant's gazelle, and warthog. Besides, Hadar, where a 4.4 million years old humanoid is recently discovered is found in this state.

K. **The Regional State of Benishangul Gumuz**

The State of Benishangul Gumuz comprises 3 administrative zones, made up of 19 woredas out of which two are special woredas (Pawi and Tongo) and 33 Kebeles. Major Tourist Attraction Sites of Benihangul-Gumuz are: Famatsere Mountain, Water Falls and Rivers, Assosa museum, various wild life's like lions, foxes, hynas, and finally, Archaeological Sites of the Region are the most common one.

2.8. Types of Visitors

The World Travel Organization (WTO) in a conference held in 1963 introduced and defined the term 'Visitor' as: 'Any person visiting a country other than that in which he has his usual place of residence for any reason other than being interested in an occupation remunerated from within the country visited. As per WTO two types of visitors are identified.

Domestic Visitor-A person who travels within the country he is residing in, outside the place of his usual environment for a period not exceeding 12 months. These are visitors that have Ethiopian nationality or live in Ethiopia and participate in different attraction areas for visiting purpose.

International Visitor –A person who travels to a country other than the one in which he has his usual residence for a period not exceeding 12 months. Visitors have not Ethiopian nationality or nationality of other country. For the tourism sector, getting more international tourists are better than the domestic ones because the revenue earning from them is by far interesting than the local tourists for the advancement of the sector. Therefore, there is no restriction to visit an area for both the domestic and the local tourists, so long as they can accommodate all the necessary service including the entrance fee.

Classification of visitors

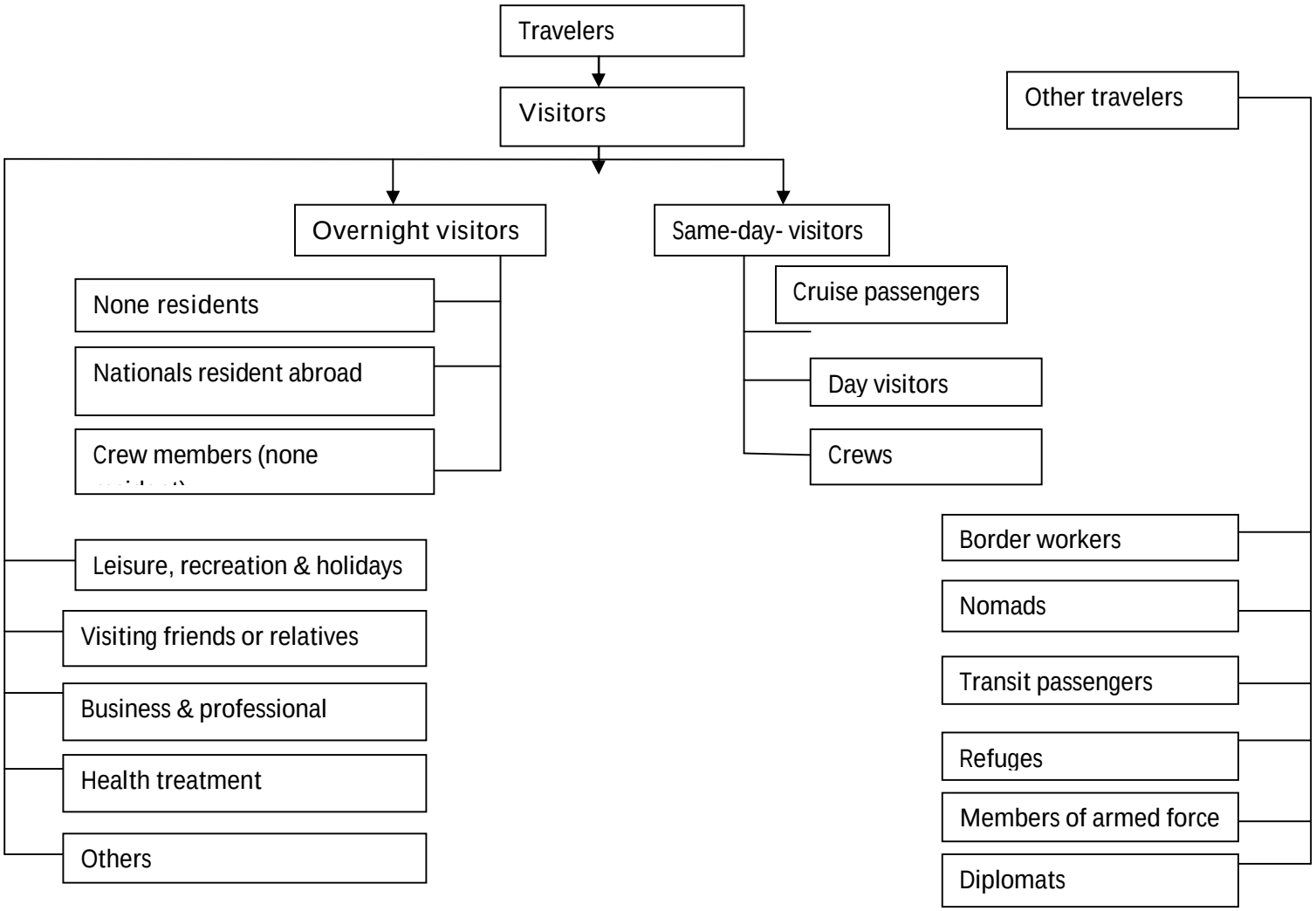


Figure 2.2 classification of visitors (source: MoCT bulletin, 2010)

CHAPTER THREE

3. Literature Review

In order to have deep understanding on the problem of this study, it is vital to review several literatures that have been conducted in the field so far. For this reason, related literature such as books, journal articles, proceeding papers, magazines, manuals and some other sources that are retrieved from the internet have been consulted so as to understand the domain knowledge, concepts, principles and methods that are important for developing recommender system and for achieving the research objective.

3.1. Introduction

The abundance of information available on the Web and in Digital Libraries, in combination with their dynamic and heterogeneous nature, has determined a rapidly increasing difficulty in finding what we want when we need it and in a manner which best meets our requirements (Pasquale, et al, 2011). Recommender systems have shown their success in many domains where information overloads exists (Santos and Boticario, 2011). Information overload refers to the fact that, for example, there are too many books, movies or songs to be able to experience all of them and make an informed decision about which ones we should read, watch or listen to. RS suggest to every user few items she might like (Avesani and Massa, 2007).

According to (Toffler, 1981), Information overload is a description given to the phenomenon where so much information is taken in by the human brain that it becomes nearly impossible to process it. The reason that information overload has become so prevalent in today's world is explained by the complexities of the communications systems available to human beings.

In a world where the number of choices can be overwhelming, recommender Systems help users to find and evaluate items of interests. The advent of the World Wide Web and concomitant increase in information which is available online has caused information overload and ignited research in recommender systems. By selecting a subset of items from a universal set based on user preferences, recommender systems attempt to reduce information overload and retain customers (Perugini, et al, 2003).

3.2. Definition and Basic concept of Recommender system

Recommender systems were originally defined as ones in which 'people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients'. It can be defined as 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. So it can be said that, the recommender system is a mapping between users and items involving a value of interest (Burke, 2002).

Recommender system is a part of a web-based application that uses data about users and their behavior to provide them with items which are the most relevant to them. The recommended items are typically things that are answerable to user taste, like books, music, news, etc. With the increase of the Internet usage and the available huge data, recommendations became a part of life (Yibeltal, 2013). No matter what the domain is, a huge amount of information is online and it becomes a difficult task to select items that are necessary. Recommender systems try to overcome this challenge and aim to map people with the correct items.

Recommender systems enable people to share their opinions and benefit from each other's experience. They were initially developed to support web users in their decision-making in daily life situations in terms of pre-selecting information that might be of interest to them, where they confronted situations without sufficient experience in the available alternatives (Santos and Boticario, 2011).

According to Yibeltal, (2013), the two basic entities which appear in any Recommender System are the user / customer and the item / product. A user is a person who utilizes the Recommender System providing his opinion about various items and receives recommendations about new items from the system.

Recommendation systems have arisen to provide convenient suggestions to the 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 (Fabiana, et al, 2003). 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 MovieLens tries to guide users to reach the movies they might like.

The aim of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them. Suggestions for books on Amazon, or movies on Netflix, are real world examples of the operation of industry-strength recommender systems.

The principal objective of recommender systems is that of complexity reduction for the human being, sifting through very large sets of information and selecting those pieces that are relevant for the active user (Fabiana, et al, 2003). Moreover, recommender systems apply personalization techniques, considering that different users have different preferences and different information needs, so the goal of Recommender Systems is to generate suggestions about new items or to predict the utility of a specific item for a particular user. In both cases the process is based on the input provided, which is related to the preferences of that user. For instance, supposing the domain of book recommendations, historians are supposedly more interested in medieval prose (Yibeltal, 2013).

3.3. 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 (Terveen and Hill, 2001). The architecture of recommender system is represented in figure 3.1

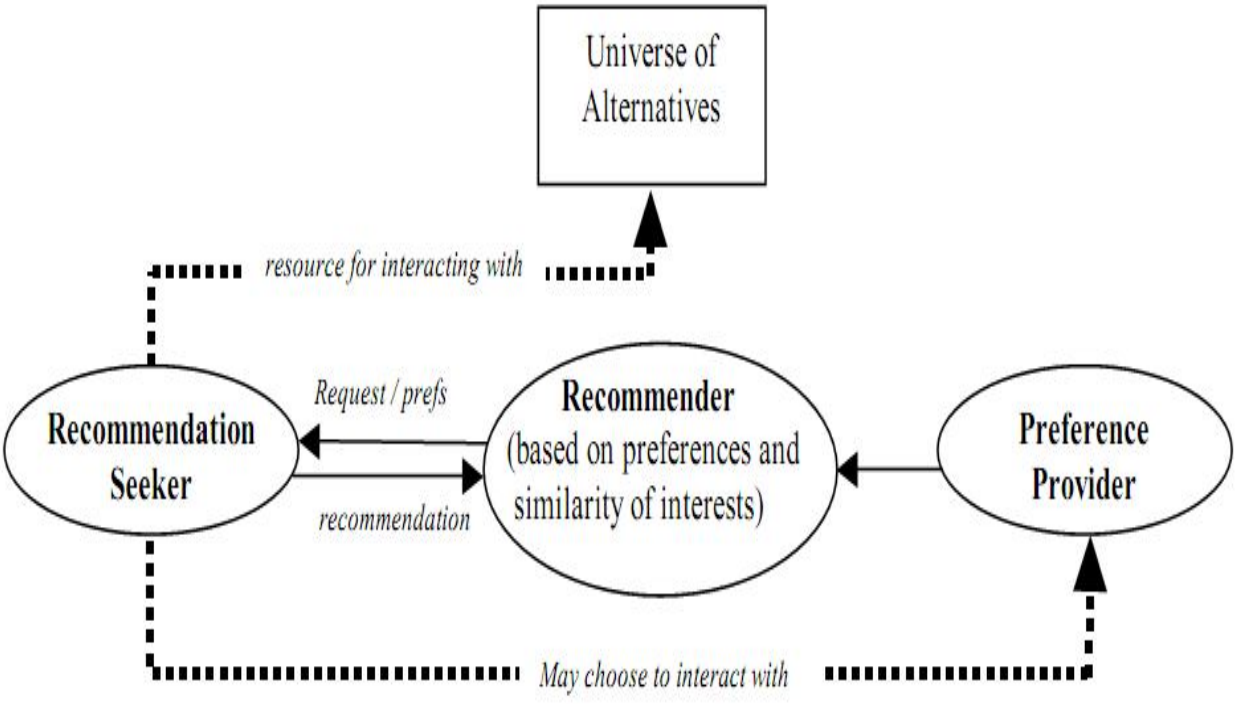


Figure 3.1 Architecture of recommender system (source: Terveen and Hill, 2001).

As shown in the above figure, the recommendation seeker (Tourists) 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 Tourist and existing case stored in the case base. Finally, the system will provide best similar cases to the tourist from the existing cases.

3.4. Recommendation techniques

Previous work of Burke, (2007) distinguished four different classes of recommendation techniques based on knowledge source, as shown in figure 3.2.

- **Collaborative**: The system generates recommendations using only information about rating profiles for different users. Collaborative systems locate peer users with a rating history similar to the current user and generate recommendations using this neighborhood.
- **Content-based:** The system generates recommendations from two sources: the features associated with products and the ratings that a user has given them. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on product features.
- **Demographic**: A demographic recommender provides recommendations based on a demographic profile of the user. Recommended products can be produced for different demographic niches, by combining the ratings of users in those niches.
- **Knowledge-based:** A knowledge-based recommender suggests products based on inferences about a user’s needs and preferences. This knowledge will sometimes contain explicit functional knowledge about how certain product features meet user needs.

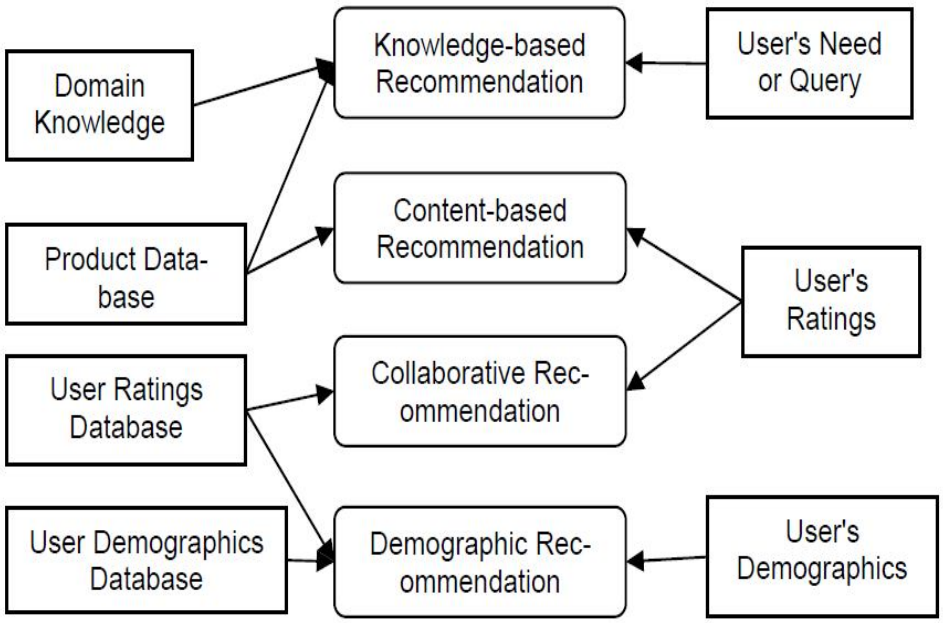


Figure 3.2 Recommendation techniques and their knowledge sources (source: Burke, 2007)

3.5. Types of Recommender system

There are various categories of recommender systems based on how recommendations are made (Burke, 2002). The different types of recommender systems are stated as follows:

- **Knowledge based recommender system:** attempts to suggest objects based on inferences about a user's needs and preferences. A system in which it depends either on explicit domain knowledge about the items or knowledge about the users to derive relevant recommendation.
- **Collaborative systems:** aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons.

It identifies users whose preferences are similar to those of the given user and recommend items they have liked.

- **Demographic recommender systems:** aims to categorize the user based on personal attributes and make recommendations based on demographic classes.
- **Content based recommendation:** In a content-based system, the objects of interest are defined by their associated features**.** A System in which the user is recommended items similar to those the user preferred in the past. It tries to recommend items similar to those a given user has liked in the past.
- **Utility based recommender systems:** make suggestions based on a computation of the utility of each object for the user.
- **Hybrid recommender system**: its core idea is to get advantages of strengths and alleviate drawbacks of each method, use the amalgam of content based and collaborative filtering recommender system. In other ways it means combine two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one.

Since the main concern of this research inclines towards knowledge based recommender system its details are discussed here under.

3.5.1. Knowledge based recommender system

As it is already mentioned above, since this research focuses on knowledge based (case based) recommender system, the detail of such system is stated bellow.

A knowledge based recommender reasons about the fit between a user's need and the features of available products and it uses knowledge about users and products to pursue knowledge based approach to generate a recommendation, reasoning about what products meet the user's requirements. The Personal Logic recommender system offers a dialog that effectively walks the user down a discrimination tree of product features (Burke, 2002). It depends either on explicit domain knowledge about the items or knowledge about the users to derive relevant recommendation. This system works based on the principle "tell me what fits based on my needs" (Biazen 2013).

As stated by Burke, (2002), there are three types of knowledge that are involved in such a system:

- ✓ **Catalog knowledge:** Knowledge about the objects being recommended and their features.
- ✓ **Functional knowledge:** The system must be able to map between the user ' s needs and the object that might satisfy those needs.
- ✓ **User knowledge:** To provide good recommendations, the system must have some knowledge about the user. This might take the form of general demographic information or specific information about the need for which a recommendation is sought.

3.5.1.1. Knowledge based system (case based reasoning system)

Knowledge is the information an expert system must have to behave intelligently. It also includes facts about the real world entities and the relationship between them. A **knowledge base** (**KB**) is a technology used to store complex structured and unstructured information used by a computer system. The initial use of the term was in connection with expert systems which were the first knowledge based systems. It is also a set of facts and heuristic (rule of thumb) about the expert system domain. Whereas A **Knowledge-based system** (**KBS**) is a computer program that reasons and uses a knowledge base to solve complex problems. The term is broad and is used to refer to many different kinds of systems (Hienes, 1983). In artificial intelligence, an **expert system** is a computer system that emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning about knowledge, represented primarily as IF-THEN rules rather than through conventional procedural code.

KBS is also one of the major family members of the AI group and it can act as an expert on demand without wasting time, anytime and anywhere. It can save money by leveraging human expert, allowing users to function at higher level and promote consistency of work. In fact a KBS is a computer based system which uses and generates knowledge from domain experts (Akerkar, et al, 2010). They are intended to perform tasks which require some specialized knowledge and reasoning. Knowledge Base systems are often called expert systems because the problems in their application domain are usually solved by human experts. For example medical diagnosis is usually performed by a doctor.

3.5.1.1.1. Architecture of Knowledge based system

KBS incorporates the following major components to accomplish its mission which is established for. These parts are discussed below both diagrammatically and theoretically.

- The knowledge base
- The inference engine
- The user interface
- The explainer and the knowledge acquisition module

Figure 3.3 below shows the building blocks of knowledge based system architecture adapted from: Hienes, (1983).

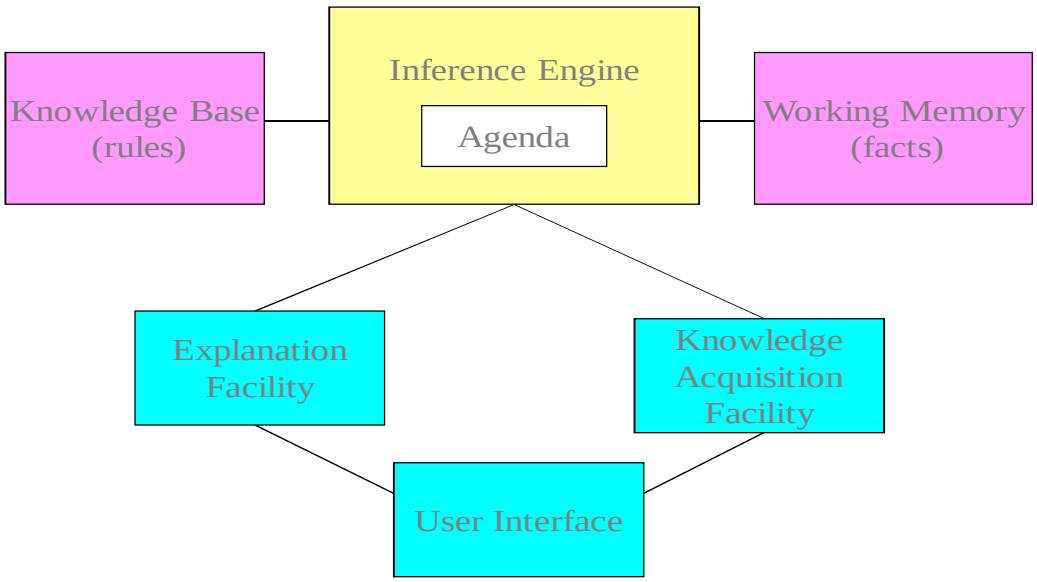


Figure 3.3: Architecture of knowledge based system.

The knowledge to which the Knowledge Base system has access, is stored in the Knowledge Base, (hence the name). The Inference Engine is the part of a Knowledge Base system which is responsible for using its knowledge in a productive way. The Knowledge Base system's reasoning mechanisms are built into the Inference Engine. Most Knowledge Base systems

employ deductive reasoning mechanisms. The Knowledge Base system communicates with the user through the User Interface.

In many applications the Knowledge Base system is required to explain its reasoning to the user. This is particularly true in situations such as the identification of chemical structures where new results must be verified. The Explainer is that part of the Expert System which provides explanation and verification. The explanation module provides a brief description to the user why the system arrived at a certain conclusion. The Knowledge Acquisition Modules help with acquiring the systems knowledge. Mostly they help to encode the knowledge from a high level format into a computer usable representation.

3.6. Knowledge Engineering

The process of acquiring knowledge from experts and building a knowledge base is called **knowledge engineering.** It involves the cooperation of human experts in the domain working with the knowledge engineer to codify and make explicit the rules (or other procedures) that a human expert uses to solve real problems (Vignette, 2004).

As stated by vignette, (2004), Knowledge engineering can be viewed from two perspectives: narrow and broad. According to the narrow perspective, knowledge engineering deals with knowledge acquisition, representation, validation, inferencing, explanation, and maintenance. Alternatively; according to the broad perspective, the term describes the entire process of developing and maintaining intelligent systems. The knowledge possessed by human experts is often unstructured and not explicitly expressed. A major goal of knowledge engineering is to help experts articulate what they know and document the knowledge in a reusable form.

3.6.1. The Knowledge-Engineering Process

The process of knowledge engineering involves *knowledge acquisition* which includes knowledge elicitation and other activities such as knowledge explication and conceptual modeling as well as the coding of the resulting knowledge, the design of a usable interface,

and the testing and evaluation of this interface (Cooke, 1994). Thus, knowledge elicitation is a sub process of knowledge acquisition, which is itself a sub process of knowledge engineering.

The knowledge-engineering process includes five major activities for the development of knowledge base system. These are: knowledge acquisition, knowledge representation, knowledge validation, inferencing and finally Explanation & justification.

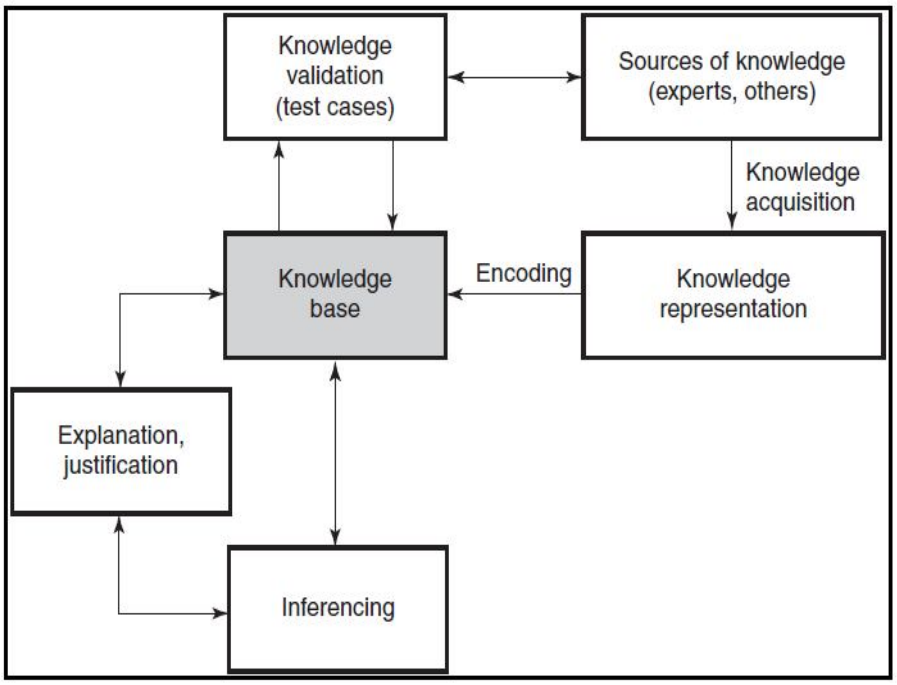


Figure 3.4 process of knowledge Engineering (source: Ranjan, et al, 2006).

The above figure shows the process of knowledge engineering and the relationships among the knowledge engineering activities. Knowledge engineers interact with human experts or collect documented knowledge from other sources in the knowledge acquisition stage. The acquired knowledge is then coded into a representation scheme to create a knowledge base. The knowledge engineer can collaborate with human experts or use test cases to verify and validate the knowledge base. The validated knowledge can be used in a knowledge-based system to solve new problems via machine inference and to explain the generated recommendation.

3.6.2. Knowledge Acquisition.

Typically knowledge acquisition is the process of building a conceptual representation of i) the overall task structure (model of the problem), ii) the expert cognitive behavior and iii) the phenomenology of the entities defining the domain. The knowledge engineer is usually concerned with acquiring knowledge about the domain and also (from the end-user) about the expected *modality* of the system (Wielinga and Breuker, 1985).

Motta, et al, (2009), defined it as the combined activity of eliciting, interpreting and organizing the knowledge acquired from the expert is called 'knowledge acquisition', and is often described as a lengthy and painful process.

Knowledge acquisition involves the acquisition of knowledge from human experts, books, documents, sensors, or computer files. The knowledge may be specific to the problem domain or to the problem-solving procedures, it may be general knowledge (e.g., knowledge about business), or it may be metaknowledge (knowledge about knowledge). (By *metaknowledge*, we mean information about how experts use their knowledge to solve problems and about problem-solving procedures in general). 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 (Ranjan, et al, 2006).

The process of knowledge acquisition includes various methods such as interview, questionnaire, record reviews and observation to acquire factual and explicit knowledge (Wielinga and Breuker, 1985). 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.

The term knowledge acquisition and knowledge elicitation have been used interchangeably in the field of artificial intelligence (AI) literature. The acquired knowledge can be specific to the problem domain, it can be general or it is meta-knowledge (knowledge about knowledge). Knowledge acquisition is the first step and time consuming task in the development of knowledge based system (Tagel, 2013). Although, KA is the most significant step in the development of knowledge base system, it is one of the most difficult and error-prone tasks that knowledge engineer does while building a knowledge-based system. Thus it needs great care, patience and attention.

Therefore, effective elicitation techniques facilitate to acquire relevant knowledge form domain experts (Tehrani, et al, 2009). KE techniques can be grouped in to different types. The well known once is direct and indirect categories. The way of categorization depends on how knowledge engineer directly obtains information from the domain expert (Wang, et al, 2011). Besides being grouped into direct and indirect categories, KE methods can also be grouped (to some extent) by the type of knowledge obtained. For example, many of the indirect KE methods are best at obtaining classification knowledge while direct methods are more suited for obtaining procedural knowledge. This does not, however, mean that the techniques cannot be used for other knowledge types. Since some designers may not be able to directly express how they perform a design task, it might be useful to use an indirect method in conjunction with a direct method to obtain this information (Cooke, 1994).

A direct method involves directly questioning a domain expert on how they do their job. In order to implement direct methods successfully, the domain expert has to reasonably articulate and willing to share his/her knowledge. However, in case of indirect methods the required knowledge is not requested directly. Instead, the result of the knowledge elicitation session must be analyzed in order to extract the required knowledge. Indirect methods are thought to be more suitable when knowledge is not easily expressed by the domain expert (Wang, et al, 2011). The commonly used knowledge acquisition techniques are discusses as follows:

A. Observation: Knowledge elicitation often begins with observations of task performance within the domain of interest. It can provide a global impression of the domain, and can help to generate an initial conceptualization of the domain. Observations can occur in the natural setting, thus providing initial glimpses of actual behavior that can be used for later development of contrived tasks and other materials for more structured knowledge elicitation methods. However, there are some tasks that cannot be observed in the natural settings (e.g., flying a one-seater aircraft) and in these Cases it may be necessary to observe performance in a simulated context or through use of a contrived task (Cooke, 1994).

B. Interview: The most direct way to find out what someone knows is to ask them. An interview technique is the process of interacting with domain expert on how they perform their tasks based on their expertise. Knowledge acquired through direct elicitation methods are procedural knowledge. Based on its structure, interview can be classified into structured, semi structured and unstructured interview (Ranjan, et al (2006). Thus, efficient and effective interview techniques largely depend on the ability of knowledge engineer to articulate their implicit knowledge. Because every interview is different in very specific ways and it is difficult to provide comprehensive guidelines for the entire interview process. Therefore, interpersonal communication and analytic skills of knowledge engineer is very important (Wang, et al, 2011).

C. Process Tracing: it involves the collection of sequential behavioral events and the analysis of the resulting event protocols so that inferences can be made about underlying cognitive processes. Thus, these methods are most often used to elicit procedural information, such as conditional rules used in decision making, or the order to which various cues are attended. The popular "think-aloud" technique in which verbal reports associated with task performance are collected and analyzed using protocol analysis is one variation on this general theme (Cooke, 1994).

D. Document Analysis: Document analysis involves gathering information from existing documentation. It may or may not involve interaction with a human expert to confirm or add to this information**.** This technique is used to collect relevant knowledge from the existed documents of different format. These documents include professional literature,

brochures, manuals, guidelines, employee handbooks, reports, glossaries, course texts, and other relevant materials (Burge, 1997).

Knowledge acquisition is not an easy task. It includes identifying the knowledge, representing the knowledge in a proper format, structuring the knowledge, and transferring the knowledge to a machine. The process of knowledge acquisition can be greatly influenced by the roles of the three major participants: the knowledge engineer, the expert, and the end user (Breuker & Wielinga, 1985).

3.6.2.1. Difficulties in knowledge acquisition

Acquiring knowledge from experts is not an easy task. The following are some factors that add to the complexity of knowledge acquisition from experts and its transfer to a computer:

- ✓ Experts may not know how to articulate their knowledge or may be unable to do so.
- ✓ Experts may lack time or may be unwilling to cooperate.
- ✓ Testing and refining knowledge is complicated.
- ✓ Methods for knowledge elicitation may be poorly defined.
- ✓ System builders tend to collect knowledge from one source, but the relevant knowledge may be scattered across several sources.
- ✓ Builders may attempt to collect documented knowledge rather than use experts. The knowledge collected may be incomplete.
- ✓ It is difficult to recognize specific knowledge when it is mixed up with irrelevant data.
- ✓ Experts may change their behavior when they are observed or interviewed.

3.6.3. Knowledge Representation

The next task after knowledge acquisition is representing the elicited knowledge in such a way that the computer can understood it. It refers to expressing knowledge explicitly in a computer-tractable way such that the agent can reason out (Davis, et al, 1993). 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 knowledge in the knowledge base.

3.6.4. Knowledge validation

Knowledge validation (or verification) involves validating and verifying the knowledge (e.g., by using test cases) until its quality is acceptable. Testing results are usually shown to a domain expert(s) to verify the accuracy of the ES. In broad sense validation is the part of evaluation that deals with the performance of the system (e.g., as it compares to the expert's). Simply stated, validation is building the right system (i.e., substantiating that a system performs with an acceptable level of accuracy). Whereas verification is building the system right or substantiating that the system is correctly implemented to its specifications.

3.6.5. Inference Engine

Inference Engine is the problem processing part of the expert system. It is the control structure of ES that allows the Expert to use search strategies to test different hypothesis to arrive at Expert system conclusions.

3.6.6. Explanation Module

An **explanation** is an attempt by an ES to clarify its reasoning, recommendations, or other actions (e.g., asking a question). The part of an ES that provides explanations is called an **explanation facility** (or **justifier**).This step involves the design and programming of an explanation capability (e.g., programming the ability to answer questions such as why a specific piece of information is needed by the computer or how a certain conclusion was derived by the computer).

3.6.7. Techniques of knowledge based Reasoning

When the knowledge representation in the knowledge base is completed, or is at least at a sufficiently high level of accuracy, it is ready to be used. There are various techniques of knowledge based reasoning. Among these the well known once are the following. Ontology based reasoning, semantic network, neural network, fuzzy logic, case based reasoning and

rule based reasoning. For the purpose of this research work, case based and rule based reasoning approach are discusses as follows.

3.6.8. Case Based Reasoning

A case can be defined as a contextualized piece of knowledge which records an episode where a problem or problem situation was totally or partially solved. A case typically represents the description of a problem situation together with the experiences gathered during the solution of the problem situation. It may also contain other items such as the effects of the solution applied or a justification for the solution and explanation or be enriched by an administrative part (including e.g., a case number) (Wangenheim, 2000).

In CBR terminology, a *case* usually denotes a *problem situation*. A previously experienced situation, which has been captured and learned in a way that it can be reused in the solving of future problems, is referred to as a past case, previous case, stored case, or retained case. Correspondingly, a new case or unsolved case is the description of a new problem to be solved. Case-based reasoning is - in effect - a cyclic and integrated process of solving a problem, learning from this experience, solving a new problem, etc (Aamodt and Plaza, 1994).

A case-based reasoning (CBR) system is a problem solver that uses the recall of examples as the fundamental problem-solving process. It also contains a number of different knowledge containers like the case base, the vocabulary in which cases are described, the similarity measure used to compare cases, and, if necessary, the knowledge needed to transform recalled solutions. A case-based recommender system is one that treats the objects to be recommended as cases, and employs CBR techniques to locate them (Burke, 2002).Case-based reasoning relates to a reasoning process based on recalling a related previous experience (a memory of stored cases recording specific prior episodes) rather than reasoning based on generalized rules (Ethiopia, 2002).

It can also means using old experiences to understand and solve new problems. In case-based reasoning, a reasoner remembers a previous situation similar to the current one and uses that

to solve the new problem. Case based reasoning can mean adapting old solutions to meet new demands; using old cases to explain new situations; using old cases to critique new solutions; or reasoning from precedents to interpret a new situation (much like lawyers do) or create an equitable solution to a new problem (much like labor mediators do)(Kolodner, 1992).

CBR is a methodology for solving problems by utilizing previous experiences. It involves retaining a memory of previous problems and their solutions and, by referencing these, solves new problems. Generally[1], a case-based reasoner will be presented with a problem. It may be presented by either a user or another program or system. The case-based reasoner then searches its memory of past cases (the case base) and attempts to find a case that has the same problem specification as the current case. If the reasoner cannot find an identical case in its case base, it will attempt to find the case or cases in the case base that most closely match the current query case (Main, 2001).

Bergmann (1998) also noted that case based reasoning is a methodology to model human reasoning and thinking as well as for building intelligent computer systems. Case based reasoning system stores knowledge in four different knowledge containers. The knowledge containers are: Vocabulary (used features), Case base, Similarity assessment, Solution adaptation.

According to Wangenheim, (2000), CBR can be an approach to develop knowledge-based systems that are able to retrieve and reuse solutions that have worked for similar situations in the past. It is also a problem solving paradigm that in many respects fundamentally differs from other Artificial Intelligence approaches. Instead of relying solely on general knowledge of a problem domain or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced, concrete problem solutions. A new problem is solved by finding a

similar past case and reusing it in the new problem situation (as it is shown in figure 3.5) below.

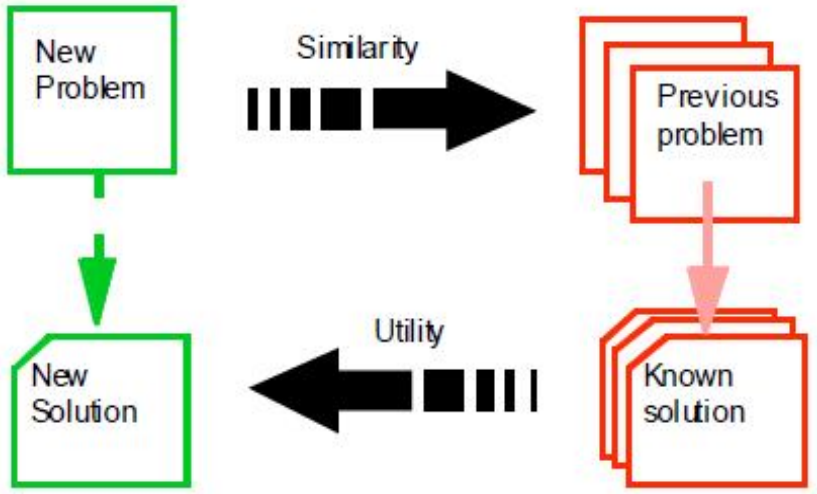


Figure 3.5 basic idea of CBR approach (source: Wangenheim, 2000).

A second important difference is that CBR is an approach to incremental, sustained learning, since a new experience is retained each time a problem has been solved, making it immediately available for future problems.

Case-based reasoning systems are an alternative in many situations to rule-based systems. In many domains and processes, referring to cases as a means of reasoning can be an advantage due to the nature of this type of problem solving. One of the most time consuming aspects when developing a rule-based system is the knowledge acquisition task. Acquiring domain specific information and converting it into some formal representation can be a huge task and, in some situations, especially less understood domains the formalization of the knowledge cannot be done at all. Case-based systems usually require significantly less knowledge acquisition as it involves collecting a set of past experiences without the added necessity of extracting a formal domain model from these cases. In many domains, there are insufficient

cases to extract a domain model. There is another benefit to CBR, which is that a system can be created with a small, or limited, amount of experience and incrementally developed, adding more cases to the case base as they become available (Mani, 2001). In case based reasoning system the data is represented in form of cases and each case consists of a *problem* and the affiliated *solutions*. Each problem is described by attributes, which are mostly represented as linear vectors and the attributes of the problems are then called *symptoms* while the solution is the *diagnosis (*Haffner, et al, 2000).

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 case(s) 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 (Ethiopia, 2002).

In the situation where a previous identical case is retrieved, presuming its solution was successful, it can be returned as the current problem's solution. In the more likely case that the retrieved case is not identical to the current case, an adaptation phase occurs. In adaptation, the differences between the current case and the retrieved case must first be identified and then the solution associated with the retrieved case modified taking into account these differences. The solution returned in response to the current problem specification may then be tried in the appropriate domain setting. The structure of a case-based reasoning system therefore is usually devised in a manner that reflects these separate stages (Main et al, 2001).

At the highest level a case-based reasoning (CBR) system can be thought of as a black box (see Figure 2.7 bellow) that incorporates the reasoning mechanism and the external facets:

- ✓ The input specification, (or problem case)
- ✓ The output suggested solution
- ✓ The memory of past cases that are referenced by the reasoning mechanism.

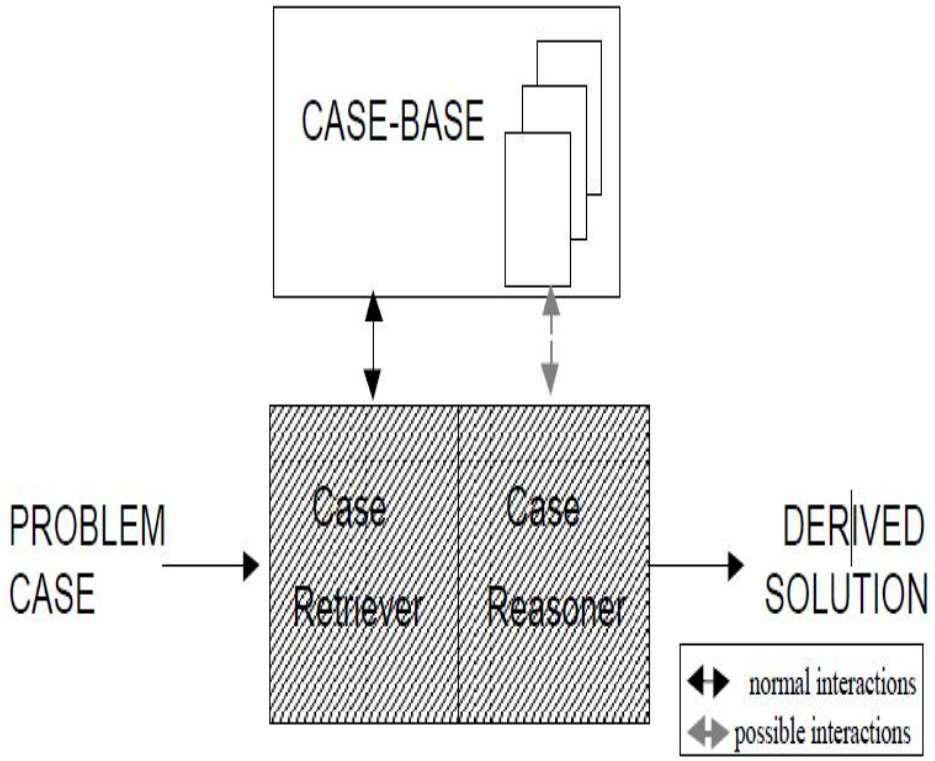


Figure 3.6 the major components of CBR system (adapted from: Mani, 2001).

As it is shown in the above figure, in most CBR systems, the case-based reasoning 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 uses the retrieved cases to find a solution to the given problem description. This reasoning generally involves both determining the differences between the retrieved cases and the current query case; and modifying the retrieved solution appropriately, reflecting these differences. This reasoning part itself may or may not retrieve further cases or portions of cases from the case base.

3.7. The CBR cycle

At the highest level of generality, a general CBR cycle may be described by the following four Processes (Aamodt, and plaza, 1994). This process comprises the four Res:

1) RETRIEVE the most similar case or cases.

2) REUSE the information and knowledge in that case to solve the problem.

3) REVISE the proposed solution.

4) RETAIN the parts of this experience likely to be useful for future problem solving.

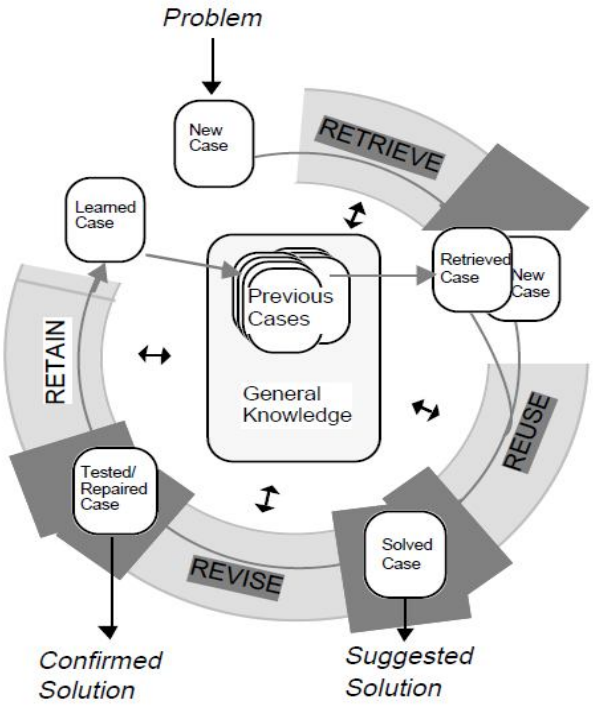


Figure 3.7 The CBR Cycle (source: Aamodt and plaza, 1994)

As shown (top of figure 3.7), an initial description of a problem defines a new case. This new case is used to RETRIEVE a case from the collection of previous cases. The retrieved case is combined with the new case through REUSE into a solved case, i.e. a proposed solution to the initial problem. Through the REVISE process this solution is tested for success, e.g. through application to the real world environment or evaluation by an expert, and repaired if fails.

During RETAIN, useful experience is retained for future reuse, and the case base is updated by a new learned case, or by modification of some existing cases.

Retrieval is the task that involves retrieving a case from the collection of previously solved cases. The retrieved case is combined with the new case for later reuse into a solved case. Revise is a process that tests the success of a solution by applying into a real world environment, if repair is failed. When useful experience is retained the case is updated by a new learned case (Aadmot and Plaza, 1994).

A new problem is solved by *retrieving* one or more previously experienced cases, *reusing* the case in one way or another, *revising* the solution based on reusing a previous case, and *retaining* the new experience by incorporating it into the existing knowledge-base (case-base). Cased based reasoning process generally involves both determining the differences between the retrieved cases and the current query case. It also involves modifying the retrieved solution to appropriately reflect these differences (Main et al, 2001).

3.8. CBR Techniques

As stated by Aamodt & Plaza, (1994), CBR has various techniques. The most common and famous once are the following: Case representation, Indexing, Storage, Retrieval and Case adaptation.

3.9. Benefits of CBR

As discussed by Ejigu, (2012).The main advantages of case-based representations are as follows:

- Ability to express specialized knowledge: This feature of cases among other advantages circumvents interpretation problems suffered by rules (due to their generality).

- Naturalness of representation: Cases are a simple knowledge representation method and very comprehensible to the user.
- Modularity: Each case is a discrete, independent knowledge unit that can be inserted into or removed from the case base without any problem.
- Easy knowledge acquisition: Knowledge acquisition in case-based representations is not usually a problem, due to the fact that cases are available in most application domains. However, there are domains where they are not.
- Self-updatability: Knowledge in the form of new cases faced during real-time operation can be incorporated into the case base extending the effectiveness of the system. This self-updatability also facilitates the maintenance of the case base.
- Handling unexpected or missing inputs: A case-based system can handle unexpected cases not recorded in the system or missing input values by assessing their similarity to stored cases and reusing relevant cases.

3.10. Limitations of CBR

However, as noted by Ethiopia (2002), case-based representations are suffered from the following problems as:

- Inability to express general knowledge: Cases by their nature, express specialized knowledge. So, they cannot express general knowledge. This is a disadvantage compared to rule-based systems;
- Knowledge acquisition problems: Although knowledge acquisition is not a problem when a sufficient number of cases are available in a domain, various knowledge acquisition problems may arise when dealing with domains where cases either are unavailable or are available in a limited (insufficient) amount;
- Inference efficiency problems: The efficiency of the inference process in CBR is not always as good as desired. Efficiency problems involve two main aspects: case retrieval and adaptation;

- Provision of explanations: Some kind of explanation can be provided for the conclusions reached, but not in a straightforward manner as in rule-based systems. It is difficult to explain all reasoning steps.

3.11. The comparison of Rule based & case based reasoning

Rule based reasoning: Symbolic rules are one of the most popular knowledge representation and reasoning methods. Their popularity stems mainly from their naturalness, which facilitates comprehension of the represented knowledge.

The basic form of a rule is the following:

If <conditions>

Then <conclusion>

Where <conditions> represents the conditions of a rule, whereas <conclusion> represents its conclusion. The conditions of a rule are connected between each other with logical connectives such as AND, OR, NOT etc., thus forming a logical function. When sufficient conditions of a rule are satisfied, the conclusion is derived and the rule is said to fire (or trigger). Rules represent general knowledge regarding a domain (Prentza and Hatzilygeroudi, 2007).

Rules are suitable to represent general knowledge, whereas cases are suitable for representing specific situations. Rules in a rule based system have the abilities to represent experiential knowledge acquired from experts in a direct fashion. Cases are capable of representing specific historical knowledge. The problem here is that it is difficult to acquire complete and perfect knowledge in a complex domain. Cases are natural and easy to obtain. They can be collected from the historical record, repair logs or other sources (Prentza and Hatzilygeroudi, 2007). CBR uses partial matching to draw a conclusion. If some of the given problem descriptions match with a given case, then the case is applicable to the proposed solution. It also tries to handle novel problems by referring previously solved cases. Rule based reasoning uses perfect matching to apply a rule for a given problem. It doesn't handle

missing information and unexpected data values (Kolonder, 1992). A case-based system can handle unexpected cases not recorded in the system or missing input values by assessing their similarity to stored cases and reusing relevant cases. It is not possible to draw conclusions from rules when there are missing values in the input data.

Thus, using the combination of both approaches, 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 (Prentza and Hatzilygeroudi, 2007).

3.12. CBR system performance evaluation methods

Evaluation of knowledge base system includes both system performance (statistical analysis) and user acceptance (Getachew, 2012). The statistical analysis for CBR can be conducted for both retrieval and reuse process. The first task of CBR is to retrieve cases that are relevant to the new case (Aamodt & Plaza, 1994). 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 (McSherry, 2001). 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 (McSherry, 2001). Whereas precision is the ratio of the number of relevant cases returned to the total number of cases for a give new case (Junker et al, 1999; McSherry, 2001).

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

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

Only system performance evaluation based on statistical analysis does not assure the applicability of the system in the real life. Even though system that achieves better system performance statistically, it may not be comfortable to the user in solving the particular problem (Getachew, 2012). As a result of this user acceptance is conducted to assess the applicability of the system for the real life.

3.13. CBR development Tools

For the development of CBR, several tools are available both commercially and freely. Most of these tools are commercial and few of them are non-commercial. The following CBR tools are indicated on the paper of Watson and Marir (1994), Yibeltal (2013) and Getachew (2012).

Kate

This tool is developed by AcknoSoft (Watson & Marir, 1994) that can run on MS Windows, Mac, or SUN. Kate is made up of Kate-induction, Kate-CBR, Kate-Editor and Kate-Runtime, this tool support both kind of Nearest Neighbor and Inductive retrieval algorithm. Kate-Induction is an ID3-based induction system that supports object-oriented representation of cases. Cases can be imported from many databases and spread sheets. Induction algorithm can make use of background knowledge. In induction algorithm, retrieval using trees is extremely fast. Kate-CBR uses nearest-neighbor approach.

Eclipse – The Easy Reasoner

It is a product of Haley Enterprises. Eclipse is implemented in C by NASA. In late 1980, the former chief scientist of Inference developed a new language like Eclipse. Eclipse is available for Dos, Windows and UNIX platforms. It supports nearest neighbor and inductive retrieval (Watson & Marir, 1994).

CaseAdvisor

It is marketed by Sentenia Software at Frazier University in Canada (Watson & Marir, 1994). It is also developed by Inference‟s CBR product. This software has three parts (Watson & Marir, 1994). These are CaseAdvisor Authoring, CaseAdvisor Resolution and CaseAdvisor WebServer.

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 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. iCOLIBRI 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.

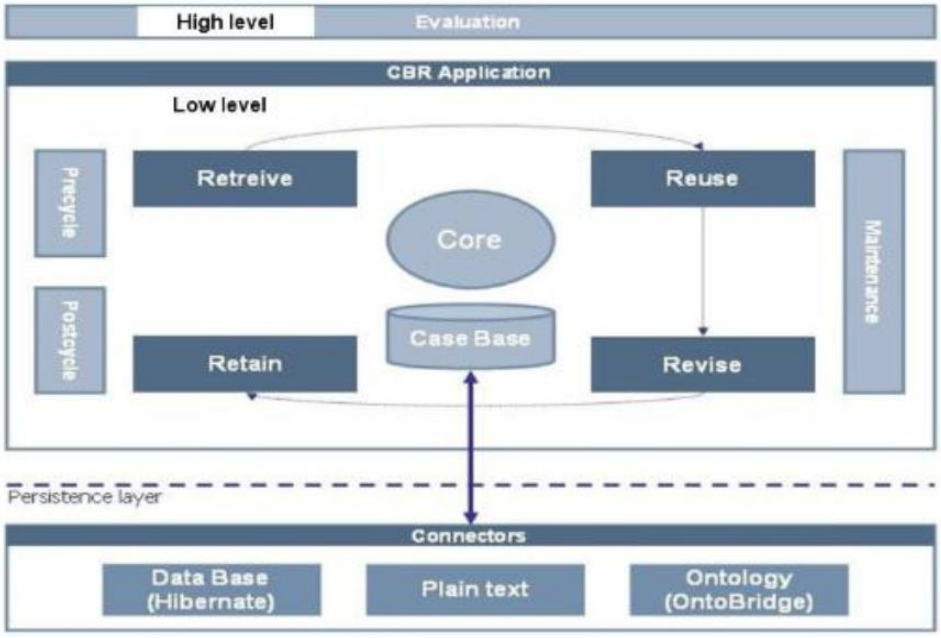


Figure 3.8 software architecture of jCOLIBRI(source: Antanassov, A. & Antonov, L. 2012).

myCBR

myCBR is one of the most popular CBR software platforms. It is a framework with certain capabilities and limitations. 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. The framework myCBR supports description of cases with various attributes: numeric, character, string, logical and class type. The templates of the cases are generated as classes or subclasses with a number of attributes, called slots. The CBR cases are objects of the class described by its attributes. Each attribute can participate in the class with its value and a weight that determines the significance of the attribute in relation to others. Attributes a weight of zero (0) is not considered when searching the case base database.

In myCBR the case and their attributes created manually or automatically. The automatic generation of attributes (slots) is done during the import procedure of the Comma Separated

Value (CSV) file. Then to each column name of the CSV file is assigned an attribute with the same name to each row of the file the new case (instance of the class) is created in the case base database. With regard to maintenance the CBR 4REs cycle phases, myCBR supports only Retrieve and Retain.

3.14. Related works on Recommender system

Even though we are often making choices without sufficient personal experience of the alternatives that are available to us in different circumstances, in our everyday life, we sometimes rely on recommendations from other people either by word of mouth, recommendation letters or on movie and book reviews to select from the huge amount of Information that is available in different places but this suggestion is not enough in this digital age (Fong and Aghai, 2009).

Since there is no local research in our country on Tourism sector of recommender system, the researcher review foreign researches on factors affecting tourist decision 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.

One of the related works is "An Automated University Admission Recommender System for Secondary School Students" by Fong and Aghai, (2009). Admission and placement of students is based on the perspective of Universities who knows little about the incoming student background but not based on the perspective of high schools who knows the detail of their students. There is value in extending the university admission process to include secondary schools (Fong and Aghai, 2009).

In this work, the author proposes a novel design of a recommender system that can provide recommendations about which universities a student should apply to, taking not only the student's secondary school scores but also other factors such as background interest and special skill into account.

In the summary of the author, education systems which do not have a standardized open exam for university admissions face the challenges of matching the right secondary school students with the right universities and field of studies and the ways that they should enter. This implies some manual processes are needed and web based recommendation system is very important for decision making. To do that, the author applied a hybrid data mining model to implement a recommender system prototype and analyze different data from secondary schools.

In addition to that, a work entitled “Recommender system for higher education” has been done to discuss the process of developing recommender system for educational institutions. The system is web based application that guides students for decision making based on their personal test. Information about course, curriculum research and facilities in the field of education is important to be available on the web. Clear information about educational activities with description enables students, partners and people to choice more efficiently and scientifically to make right decision (Satyanarayana and Rajagoplan, 2007).

The author concludes that, we can make beneficial use of artificial intelligence techniques like database design and selection, content based recommendations, user profiling, integrating groups of users with similar interests and integrating the domain knowledge and expertise. Hybrid recommendation system approach is important in educational institutions (Satyanarayana and Rajagoplan, 2007).

Bendakir and Emeur, (2003), also tried to discus about course recommendation using data mining techniques called association rule. According to the author, students often need guidance in choosing adequate courses to complete their academic degrees. Course recommender systems have been suggested in the literature as a tool to help students make informed course selections.

Students who join higher education degrees are faced with two main challenges: a myriad of courses or field of studies from which to choose, and a lack of knowledge about which courses or field of studies are relevant to follow and in what sequence. Mostly, it is according

to their friends and colleagues' manual recommendations that the majorities of them choose their field 40 of study and register for it. It would be useful to help students in finding courses of interest by the intermediary of web based recommender system (Bendakir and Emeur, 2003).

The main focus of the authors was on the effectiveness of the incorporation of data mining in course recommendation. The system is based on the following collaborative filtering algorithms: user-based and item-based. According to the author, the system can predict the usefulness of courses to a particular student based on other users' course ratings. To get accurate recommendations, one must evaluate as many courses as possible. Based on the evaluation results, the author suggests C4.5 as the best algorithm for course recommendation. The system cannot predict recommendations for students who have not taken any courses at the University. Generally, there are many recommender systems developed globally and few attempts have been done locally on the area of recommender system in different sectors which is stated bellow.

Biazen, (2013) 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 is 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. After accepting the input the system calculates similarity between existing case and new queries that are provided by the students and provides solution or recommendation by taking best cases to the new query. 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 user interface which enables students to enter their query and programming codes with the help of Java script language. 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, (2013), has also 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.

Finally Yibeltal, (2013) has done research study on application of recommender system in investment activity selection in the case of Ethiopia. The objective of the researcher was to develop Case Based Recommender System that can give recommendation on the selection of investment sectors and investment activity in Ethiopia to foreign and domestic investors. Here also the system provides recommendation to new investors based on previously solved cases and new query given by the investor. The author uses 1344 cases which are collected from successful investor as case base. In this study the author also used the same development tool with that of Biazen's work to develope the prototype of case based recommender system. 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.

Finally Based on user acceptance of prototype testing, the average performance of the system is 82% and 84% by the domain experts and investors respectively.

Similarly, the proposed knowledge based systems in this study is conducted to explore the applicability of case based recommender system for the tourism sector in the selection of tourist attraction areas and suitable visiting time. The main objective of the research study is to assign visitors in different tourist attraction areas with the appropriate time based on personal characteristics, socio economic characteristics and other factors. Therefore, the proposed case based recommender systems can assist tourism expert and tourists during their recreational program. This work will also be used as an input for further development of Tourism related works. Related works are summarized in the table below.

Author	Research title	Methods /techniques	Algorithm	Programming tool	No. of cases used	Retrieval		Accuracy
						Recall	Precision	
Henok Bekele (2011)	CBR Hypertension management	feature-value case representation	NNR	python	45	86.1%	60%	88.89 %
Getachew (2012)	CBR for anxiety disorder diagnosis.	feature-value case representation	NNR	jCOLIBRI	50	82%	71%	NE
Biazen (2013)	CBR Field of study selection	feature-value case representation	NNR	jCOLIBRI	105	85%	55%	NE
Yibeltal (2013)	CBR Investment advice system	feature-value case representation	NNR	jCOLIBRI	1344	85%	64%	87%

CHAPTER FOUR

4. EXPERIMENTAL DESIGN

In this chapter, the researcher collects tourist cases and models it by using hierarchical structure. This aids to develop the prototype using case-based reasoning. The knowledge for this study is acquired from domain experts by using interviewing and critiquing knowledge elicitation methods and from relevant documents by using document analysis technique which has been employed to purify the acquired knowledge.

As discussed by Tagel (2013), there are certain important steps that the knowledge engineer need to carry out during knowledge acquisition process. These are:

- Eliciting data and information from the domain experts.
- Interpreting the acquired information to understand human expert reasoning processes.
- Constructing model to represent the expert's knowledge.
- Repeating step I-III as the knowledge base system involves into a functional system.

There are two main steps in knowledge acquisition process that are accomplished by the knowledge engineer so as to develop knowledge-based system. These are knowledge elicitation and knowledge structuring.

4.1. Knowledge Elicitation

This involves extracting knowledge from human experts, and/or written documents to build a knowledge-based system. In this study, the knowledge required to build a knowledge-based system was elicited from both tacit and explicit sources of knowledge. Tacit knowledge is collected from eight experts in the domain area from NTO by using semi-structured interviews (the sample interview questions used are found in Appendix I). Here, the number of experts to gather the needed knowledge was determined as eight because there were about 17 domain experts at NTO and interviewing all of them is challenging and time taking for

both the experts and the researcher. And it can be the representative of them. Domain experts are chosen purposely for wide-ranging discussion using semi-structured interviews to understand the domain knowledge. These experts are essentially taking part during the study and asked to verify the rightness of the acquired knowledge. Moreover, explicit source of knowledge has been collected from the internet, manuals, research papers and journal articles, etc.

4.2. Knowledge structuring

It involves using concepts discovered during the knowledge elicitation session to build a model or representation of the domain experts. It is a process where knowledge engineer uses concepts discovered during the knowledge elicitation phase to build a model of the domain. The knowledge used for building of the knowledge-based system in this study focused on knowledge regarding the tourist cases.

The process of knowledge acquisition of this research encompasses some basic activities such as gathering the needed knowledge, analyzing that knowledge, identifying important concepts (tourist attraction areas) and finally modeling them in using hierarchical structure. The fundamental knowledge which is used to develop recommender system was acquired from various sources.

4.3. Knowledge Acquisition from Domain Experts

For this study, original tacit knowledge is collected from domain experts from NTO via interview. Since interview is one of the knowledge elicitation techniques which involve asking the domain expert on how they perform their task and become successful. To collect the required knowledge, the researcher used semi-structured interview technique. As one of the great intentions of this chapter is eliciting relevant tacit knowledge from the domain experts, eight (8) domain experts from each Tourism sectors weres selected using purposive sampling technique. As a result of this, Tourism officers from each Tourism sector have been interviewed to obtain the required knowledge on the domain area.

The domains of interview with those experts covers issues like, how the expert interact with tourists, what are the major criteria to be considered to assign Tourists in different Tourist attraction areas, and what are the possible Tourist attraction areas recommended to the Tourists and appropriate visiting time (see full interview questions in appendix I).
The researcher tries to make an extensive discussion with the experts to acquire the relevant tacit knowledge which is significant to generate the proposed case. In addition, the domain experts are actively participated throughout the research work and they are consulted to confirm the correctness of the acquired knowledge. During face to face communication, the acquired knowledge from domain experts has been recorded manually by using pen and paper sheet.

Tourism experts are devoted in providing promotional services in the institution. According to some domain experts, the investigation of visiting application starts by collecting some relevant information such as gender, travel frequency, age, annual income, and type of Visitor (nationality) i.e. whether they are domestic or foreign, attraction preference, length of stay, etc. But even if these visitor's files are collected, mostly experts assign visitors in different tourist attraction area based on their interest. During the discussion, almost all domain experts provided the same information about tourist attraction area selection advising system and they said that since there is no clear and consistent tourism guideline or criteria to assign or recommend visitors in different tourist attraction area, they simply asks visitors about their interested attraction area then gives some explanation about the selected attraction area rather than use a guideline or criteria to recommend tourists in suitable attraction areas.

Since one of the main target of this research is to identify the most important factors that has a great impact on the selection of tourist attraction area, the researcher forwarded few questions to domain experts about "what are the main attributes having effect on tourist attraction area selection?". The researcher also asks the effect of some attributes that are found from different secondary sources to get confirmation from domain experts whether it has effect or not. Based on the above questions domain expert said that, considering some of

the profile of visitors and other attributes are important to determine the type of tourist attraction area to visitors; because some tourism sectors and tourist attraction areas have unique associations with annual income, age factors, gender factor, travel frequency factor, type of visitor factor, travel purpose (see the detail in section 4.8)

At the end, domain experts conclude that, considering the above factors to assign visitors in different tourist attraction area is a very important thing to facilitate the service. Because tourists mostly choose the attraction area based on the opinion of others or their interests without considering the suitable time period of visiting, and the other factors. Hence, experienced and knowledgeable experts can identify the best attraction area to each visitor by considering different demographic characteristics of visitors, socio-economic factors and other factors to make immediate decision.

4.4. Knowledge Acquisition from Tourists

Primary sources of knowledge are also collected from visitors through interview. To elicit the required knowledge again semi-structured interview technique has been employed. Since one of the main aims of interviewing visitors is to elicit relevant tacit knowledge, about fourteen (14) visitors are selected using random sampling techniques and accordingly they have been interviewed. During the acquisition of knowledge from visitors, the researcher has got about 127 lists of tourists from NTO who come to the country within a month. The researcher also made an effort to know the time period in which these tourists can be available to make interview. Based on the information obtained from experts, fourteen visitors were selected randomly as a representative of the whole and they have been interviewed.

The domains of interview with tourists covered issues such as how to get advising services from domain experts, what are the problems in advising systems of NTO and is there any experienced Tourism experts that gives a brief advice on how to visit and where to visit (see the interview question in appendix II). Based on the questions raised, visitors responded that, there were no sufficient and experienced tourism experts that could provide advice for each

queries. Tourists further comments that, the advising system is quite different from expert to expert. Due to these reason, tourism experts mostly recommend visitors on the area which is always familiar with them. Visitors said that it must need well organized and consistent advising guidelines for tourists. Difficulty of getting appropriate advice is a critical issue for visitors since knowing the right attraction area is a key factor and also knowing the right time to visit is another key factor to consider to new tourists. Visitors tend to lose money by choosing wrong tourist attraction areas and in appropriate time.

4.5. Knowledge Acquisition from Relevant Documents

There are often several sources of explicit knowledge. Since printed information can sometimes become outdated, printed materials should never be considered as sufficient sources of information for the development of a knowledge-based application. Document analysis involves gathering knowledge from existing documentations. Hence, document analysis has been carried out to acquire explicit knowledge which is found in various secondary sources of knowledge.

For this research in order to elicit the explicit knowledge, relevant documents which are related to tourist attraction area and visiting time selection process have been reviewed. The most important documents used for this study are: Company Policy Manuals and Regulations, Reports, Memos and Guidelines, Published Books and Journal Articles, research papers, etc. Tourism Websites related to the tourism sector and tourist attraction area selection decisions especially NTO and MoCT website are also reviewed. As a result, relevant and technical knowledge were extracted and structured in a manner that is suitable for knowledge modeling and finally knowledge representation. The main data source (previous tourist case base) used for developing case based reasoning for Tourist attraction area and visiting time selection for this research is previous visitor's cases found from NTO office. In addition, list of tourism sectors, attraction areas, and legal frameworks on the amount of annual income for foreign and domestic visitors; have been collected from the manuals of Ministry of Culture and Tourism office.

The detail part of this knowledge acquired from different sources that focus on Tourism sector and tourist attraction area selection is discussed, structured and conceptually modeled in section 4.7.

4.6. Selecting an Attribute using Data Mining tools

To solve the problem on the collected data set of Tourist cases, there were a series of activities that are undertaken in this research work. The major activities are discussed as follows.

4.6.1. Data collection

Since the data set to be used for this research has been collected from NTO and MoCT, any field work to gather the data was not required. But, integration and preprocessing the dataset was held in order to suit it with the intended purpose**.** These datasets (previous tourist's data) which were collected was recorded in year 2006, 2007, 2008, and 2009. Originally, the dataset consists of 18 attributes and about 21,347 records, which include the relevant information concerning tourists.

4.6.2. Data preparation and cleaning

Preparing data is an important preprocessing step for data analysis not only as part of data warehousing but also for data mining. The main reason is that, the quality of the input data strongly influences the quality of the analysis results. Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing. Preprocessing the data helps to fill some missing values; to detect some outliers that may put the data at risk; and to remove or correct some noisy data. In general, data preparation comprises several subtasks: selection, integration, transformation, cleaning as well as reduction and/or discretization of data. Often these tasks are very expensive. The data originally were available in a hard copy format so, encoding these data in to a computer was one of the toughest task and time consuming. The

researcher converts the data in to electronically recorded format by recording them in Ms-excel and to use in Waitko Environment for Knowledge Analysis (WEKA) software it should be converted in to comma separate value (CSV) or Attribute Relation File Format (ARFF). Then data preprocessing have been made for cleaning it using WEKA software. After that, attributes that have so many missing values can easily be detected and removed. The total dataset collected from NTO was 21,347. Of these the number of satisfied / successful visitors' case was 1079. Then the researcher converts these 1079 case in to ARFF format and provides it to WEKA software. After that, WEKA shows those attributes whose values are missed and values as normal. WEKA shows that, the attribute nationality has a missed value of 267 values and the attribute travel purpose has a missed value of 103 where as the attribute length of stay has a missed value of 94 and the remaining 615 dataset was depicted as a normal and it was used as a case base. Values that are not in Weka compatible form are modified. For instance, Weka doesn't handle values with space unless they are in single quotation. Therefore, since it is difficult to put all such values in single quotation, the researcher preferred to put them as one token by using underscores (_) to fill spaces.

There were several problems in the original data set which needs further preprocessing. The most common and observed once were stated here under:

- Attributes have so many missing values. E.g the attribute nationality has a missing value of 267, travel purpose has a missed value of 103 where as length of stay has a missed value of 94.
- In the original dataset there is an error on changing the attribute values from attribute to attribute. For instance some records have age value in gender attribute and gender value in age attribute. Like the attribute age has a value of "M" at some place "F" in another place. Whereas the attribute Gender has a value of number in some places.
- Another problem in the dataset is there is no common writing way of attribute values i.e. some attribute values are written in abbreviation format and some attribute values are written in standard form. Eg. In the attribute travel purpose, one of its value vacations was written as vac. in some places and the full name vacation in other places.

And in the attribute nationality, its value Ethiopia was written as Eth. In some places and the full name Ethiopia in another place.

- The spellings of some attribute value are different in different places. For instance, the attribute attraction preference has a value of Semen Mountain in some places and simien national park in another place.

Hence, the researcher takes time to correct and normalize the above stated problem of the dataset step by step manually, using Ms excel and Weka tool. After going through the data cleaning, the data is saved as CSV file format in which the values are saved in comma delimited form in order to create an ARFF format file. Finally the six hundred fifteen (615) data that were converted into ARFF file format and has been used for the experiment for attribute selection.

4.6.3. Selecting Best Attributes (Features)

Feature selection is a process of selecting a subset of relevant features for use in model construction from original features. The main idea of feature selection is to choose a subset of input variables by eliminating features with little or no predictive information. Feature selection can significantly improve the comprehensibility of the resulting classifier models and often build a model that generalizes better to unseen points. Attribute selection can be used to investigate which (subsets of) attributes are the most important ones. In data mining task, one can get some attribute that has little or no impact on the overall Tourism sector and tourist attraction area selection output. Feature selection reduces the number of features, removes irrelevant, redundant, or noisy data, and brings the immediate effects for applications: speeding up a data mining algorithm, improving mining performance such as predictive accuracy and result comprehensibility Therefore since there were several attributes in the recorded tourist data set, the researcher performs an attribute selection task using Weka attribute selection algorithm. There are two major parts of attribute selection methods in WEKA: these are: search methods: like ranking, best-first, forward selection, random, exhaustive, and genetic algorithmic. The other method is an evaluation method which

encompasses information gain, correlation-based, wrapper, chi-squared etc. For the purpose of this research work, the researcher used a search method of ranking and an evaluation method of information gain attribute selection. The reasons for selecting information gain evaluation method and ranking search methods are: information gain attribute evaluation evaluates the worth of an attribute by measuring the information gain with respect to the class. Information gain evaluator is used to select the best attribute at each node in the tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split. The attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node. This attribute minimizes the information needed to classify the samples in the resulting partitions and reflects the least randomness or "impurity" in these partitions.

The attribute with the highest information gain is considered as the most discriminating attribute of the set under consideration. So, an attribute that yields maximum information gain will be chosen for data set partitioning. Ranking search method is used to rank attributes by their individual evaluation from highest information gain value to lowest information gain value. The highest information gain value (i.e. ranking first) is the most important attribute for tourist attraction area and visiting time selection. Then an experiment was conducted to differentiate the most discriminating attribute and based on the experiment, using information gain evaluation methods and ranking search methods the importance of attributes in ranking order are: attraction preference, Gender, Age, Length of Stay Travel Purpose, nationality, and location of attraction area which is going to be visited**.**

Finally, after selecting attributes using attribute selection algorithm, the researcher consulted domain experts for the purpose of validating the selected attributes are important in attraction area selection decision.

4.7. Knowledge Modeling

Conceptual Modeling of domain knowledge implies capturing the static structure of information and knowledge types (Schreiber, 1999). Models are applied to express the important characteristics of real-time systems to understand in a simple way by dividing them into small parts. Models are more related with problem domain they represent (Savolainen, et al, 1995). Real-time systems are huge objects comprising of interconnected elements doing in complications as teamwork. Models assist individuals to weigh-up and know such complications by supporting them to explore every specific area of the system. They are applied in the construction process of systems to draw the architecture of the system and to simplify the exchange of information between several individuals in the group at various levels of abstraction. Individuals have several understandings of the system and models can assist them to know these understandings in a coordinated way.

After the knowledge acquisition phase is completed from previous tourist case, domain experts, and relevant documents, the next task is modeling the knowledge. As noted by Schreiber, et al (1999), the modeling process builds conceptual models of knowledge-intensive activities. During the knowledge acquisition process, the knowledge engineer will attempt to understand both the tacit and explicit form of knowledge and then use visual tools to make an exchange of views between domain experts and end-users. This exchange of views produces concepts and understandings with regard to how the acquired knowledge is applied, how judgments are made, and so on. And the knowledge engineer should build the knowledge model from the acquired exchange of views with domain experts and end-users. This helps the knowledge engineer to transfer the knowledge model into functional computer-machine programs.

According to Schreiber et al. (1999), Knowledge modeling is very significant for knowing the operational means in the development process of a knowledge-based system. It is also a vital stage of the knowledge engineering process. It can provide a means to easily understand the

source of knowledge, the inputs and outputs of knowledge, and the designation of other parameters.

Although, there are various conceptual modeling techniques, for this study hierarchical tree structure is used to model how tourist attraction area selection is performed. The reason for using this one as modeling technique is, it easily model concepts and clearly explains the concepts in the problem area. It models the knowledge in the hierarchical manner. This model starts from the main concept at the highest level of the hierarchy and other sub concepts that can affect or affected by the highest level concept put next to down ward in the hierarchy.

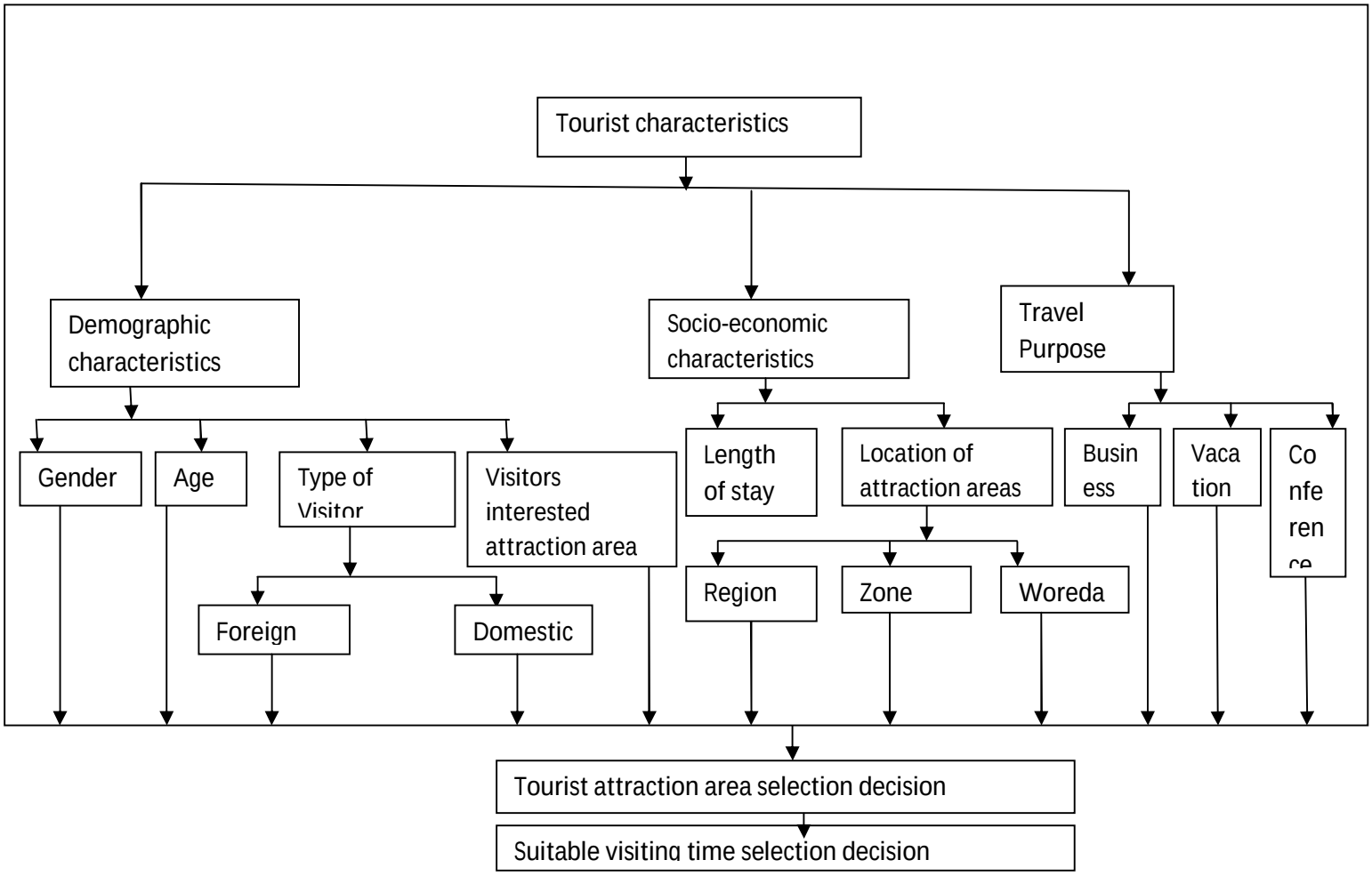


Figure 4.1 hierarchical structure of tourist attraction area selection

4.8. Factors that affects the selection of tourist attraction area decision

Everyone can be specifically differentiated on various parameters in tourist attraction area selection decision. The major factors that influence attraction area selection decisions are on the tourist's demographic characteristics such as Gender, and Age, type of tourist (nationality) and socio-economic factors such as Annual Income, and location of attraction area or environment. The detail of each factor is discussed below:

4.8.1. Gender

Gender plays a vital role in the travel decisions. Males have been noted for travelling often as a result of work associated issues or businesses, while females on the other hand are noted for travelling more for recreational reasons. This means decision on destinations to visit could be reflected in gender. Gender is related to personality, visitor‘s way of acting and their intentions to go towards attractions. Opposite views in socio- psychological intentions which represent push and pull factors that determine destination choices. These pull and push factors were analyzed on leisure travelers from a gender point of view. There is also a difference in the gender decisions as concerns push and pull factors. Females were seen to make decision associated with culture, family bonding opportunities and ego while male's decisions were related to sports, and new experiences when planning on destinations to visit (Jo¨nsson and Devonish, 2008).

On the other hand, females are more risk averse in their choices of destinations to visit and consider educational opportunities while males are adventurous or risk taking in their decisions. In addition to that, males and females of the same age group could have different views as regards leisure destination choices. In business travels especially considering hotel choices and service delivery, women prioritized security, self service and lower prices as opposed to men who valued business services and things associated with more when making decisions.

Females had stronger motivations to travel than males. They also found significant gender differences regarding travel motivations where male tourists preferred more recreation and

activity in the destination, and female tourists had a stronger relaxation and escape-based motives**.**

4.8.2. Age

The age of a tourist had a significant effect on only cultural motivations and relaxation-based motivation. For the cultural motivations factor, tourists in the oldest age categories were more likely than tourists in the youngest age category to travel to the destination based on the need ‘‘to increase their knowledge of local places’’ and ‘‘to meet local people.’’ For the relaxation-based motivations factor, tourists in the oldest age category (56 and over) were more likely to travel to the destination based on the need ‘‘to relax’’ and ‘‘to enjoy good weather,’’ compared with those in the youngest age category (18 to 35 years). Tourists in the 36 to 55 age group were more likely to travel to the destination based on the need ‘‘to be emotionally and physically refreshed,’’, compared with tourists in the youngest age category (18 to 35 years). Tourists in the youngest age category were more likely to travel to the destination based on the need ‘‘to engage in sports,’’ compared with tourists in the oldest age category (56 years and over) (Jo¨nsson and Devonish, 2008).

With respect to age differences, older tourists were more likely to travel for reasons based on cultural exploration and relaxation, whereas younger tourists were more likely to travel to engage in sports. Older tourists (who are likely to be retired and have more free time) tend to desire mental stimulation and prefer to visit countries to increase their knowledge and awareness, and learn new experiences. Younger tourists are more active and are more likely to seek a whole range of physical activities when visiting a destination (Jo¨nsson and Devonish, 2008).

Generally, Tourists who desire noisy, active, and interactive experiences in tourist destinations were more likely to be young and male, whereas older tourists tend to desire relaxation and to have a need to discover new places and things in the destination. In addition, females had stronger motivations to travel than males and male tourists preferred

more recreation and activity in the destination, and female tourists had a stronger relaxation and escape-based motives.

4.8.3. Types of Visitors

As it is discussed in chapter two and as per WTO two types of visitors are identified as one factor in the selection of attraction areas.

Domestic Visitor-A person who travels within the country he is residing in, outside the place of his usual environment for a period not exceeding 12 months. These are visitors that have Ethiopian nationality or live in Ethiopia and participate in different attraction areas for visiting purpose.

Foreign /International Visitor –A person who travels to a country other than the one in which he has his usual residence for a period not exceeding 12 months. Visitors have not Ethiopian nationality or nationality of other country. For the tourism sector, getting more international tourists are better than the domestic ones because the revenue earning from them is by far interesting than the local tourists for the advancement of the sector. Therefore, there is no restriction to visit an area for both the domestic and the local tourists, so long as they can accommodate all the necessary payments including the entrance fee.

4.8.4. Travel purpose:

Travel purpose refers to the reason that, the entrance of tourists in the country. It might be for business, visiting relatives or friends, for leisure, for education etc. Therefore, visitors should specify the motive that initiates to make a tour in a certain attraction area before they start their visiting activity because some places may not give sense for some visitors if it didn't match with his/her motive.

4.8.5. Length of stay:

It is the time period that tourists have, to visit a certain attraction. It can be for 2 weeks, a month, or a maximum of 12 months. Here visitors have to put the exact time that they have, to make a tour while visiting because if one tourist want to stay in a certain area for about a

month, that specific attraction area may not have sufficient facilities like accommodation, food and beverage to stay more and the tourist may get suffer in such a situation.

4.8.6. Location of attraction areas

As it is mentioned in chapter two, tourist attraction areas are located in various regions and its location is a critical factor for success or to be satisfied in recreational program. A good location may facilitate a struggling business to ultimately grow and creates some motivation for visitors, whereas an attraction area situated in a poor location will be at a disadvantage. Some factors to consider when establishing your attraction area location if it is manmade is think about the availability of different services like Infrastructure and facilities (such as transportation infrastructure, communications), accommodation, etc. is a key driver in strengthening the national economy and enhancing Ethiopian productivity. Infrastructure availability promotes both foreign and domestic visitors because infrastructure growths are associated with greater accessibility and reduction in transportation costs and maximization of profit (Ministry of culture and tourism, 2012).

Location of tourist attraction area includes region, zone, and woreda/city as depicted in the fig bellow.

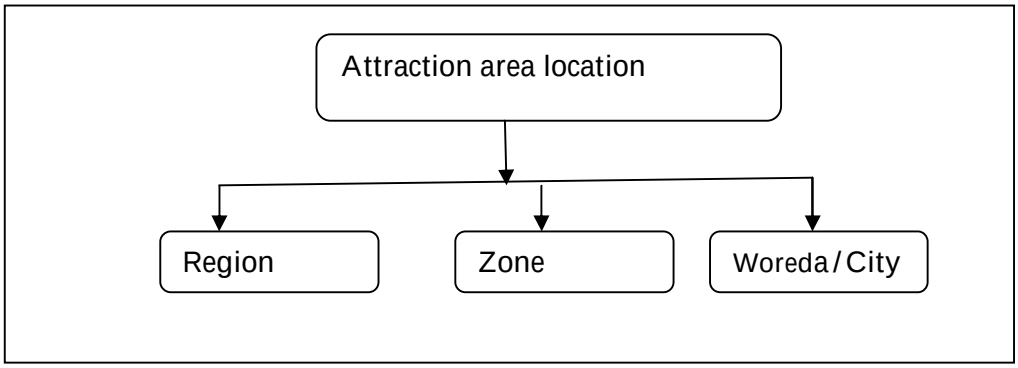


Figure 4.2 location of attraction area

4.9. The case structure of Tourist attraction area selection

The case structure of tourist attraction area selection has two basic parts. The first part is the problem (tourist's cases) descriptions or situation and the second one is the solution.

Query Description/Situation: It is the part of the case structure that consists of attributes which describes about visitors.

Solution: This part of the case structure provides the recommended attraction area, appropriate time to visit the area based on the problem descriptions (visitor's information).

Thus, for this research the researcher identify different description and solution attributes with the help of tourism expert and from recorded data set of previous tourist's cases. But, there were different challenges during identification and representation of case structure. For example, the first challenge was the attributes (the tourist's information in this case) registered in MoCT and NTO office was too many, so the researchers selects the most significant attribute that used to determine the selection of tourist attraction area using data mining attribute selection algorithm by using Weka software (see these important attributes in the table 4.1). Then the researcher selects 9 significant attributes from the total of 18 ones. The second challenge was some of previous visitors information was recorded in the form of hard copy and the researcher lost a lot of time to encode them in electronically recorded format for the development of recommender system.

The most important attributes that affects the selection of tourist attraction areas decision are listed below:

Attribute name	Parameter
Gender	Description
Age	Description
Type of visitor(nationality)	Description
Travel Purpose	Description
Length of Stay	Description
Region	Description
Zone	Description
Woreda/town	Description
Interested attraction area	Description
Recommended Visiting time	solution
Recommended attraction area	Solution
Explanation Facility	Solution

Table 4.1 the case structure for tourist attraction area selection.

The Short descriptions of attributes that are used for building the case structure are presented as follows:

Gender: is the term used to refer to a person's self-representation as male or female. The values of this attribute are either male or female.

Age: is the age of the tourist which can be labeled as number.

Type of visitor: indicates the nationality of visitor's. The value of this attribute is domestic and foreign. Domestic visitor means visitors having Ethiopian nationality and foreign means visitors having other nationality (a visitor who has not Ethiopia nationality).

Travel purpose: it is the reason that, the entrance of tourists in the country. It might be for business, visiting relatives or friends, for leisure, for education etc.

Length of stay: it is the time period that tourists have to visit a certain attraction. It can be for 2 weeks, a month, o r a maximum of 12 months.

Region: is the regional or city administration place where visitors plan to visit. 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 attraction. Since each region has its own zone, visitors select specific zones based on the selected region.

Woreda: is the woreda or city location of tourist attraction that specifically facilitates service to visitors. So visitors select a specific woreda or city location of attraction based on the selected zone.

Visitor's interest of attraction: is the process of visitor's interest of attraction that wants to visit. So visitors fill (enter) their own interest of attraction as a query.

Recommended attraction area: it is a solution and provides a recommended attraction area based on the similarity of cases.

Recommended visiting time: it is also a solution and provides recommended time to make a tour and visit a certain attraction so that visitors can easily identify the right time to visit.

Explanation Facility: is used to give explanation and description about the recommended attraction areas as well as the time period which is suitable for visiting.

CHAPTER FIVE

5. PROTOTYPE DESIGN AND IMPLEMENTATION

5.1. Introduction

In the 4th chapter of this study, the concept of knowledge acquisition and modeling parts which is the experimental design have been discussed in detail. Under this chapter, the design and implementation part of the study is given more attention and this section involves the actual development of a workable prototype CBR system for tourist attraction area selection for new visitors. Thus, having all the necessary previous visitor's cases from NTO as well as MoCT and the knowledge from various sources like from domain expert, tourists and different relevant documents, the next task is coding the knowledge into computer using appropriate and efficient knowledge representation methods. Then, after the knowledge has been represented, the next task is to develop the prototype recommender system for tourist attraction area selection. For this research, jCOLIBR 1.1 CBR frame work is used to develop the prototype and construct the case structure of recommender system. Although, there are different case retrieval algorithms like: Expert Clerk Median scoring, filtering method, and nearest neighbor, and inductive retrieval methods the researcher used nearest neighbor retrieval algorithm. This is because jCOLIBRI uses this algorithm for retrieval task. Nearest neighbor retrieval algorithm is also suitable when there are attributes which have numeric (continuous) value (Juan A. et al, 2002). The algorithm also retrieves the case which is nearest to the user`s query by measuring its similarity with the cases.

5.2. Architectural Design of CBR system for Tourist Attraction Area Selection

Figure 5.1bellow illustrates that, the framework of case based recommender system and how the prototype works for tourist attraction area selection in Ethiopia. As the new query (problem) is entered, the prototype of the system matches the new case with the solved case in the case base of the system by using similarity measurement. If relevant cases are found

within the case base, then the prototype system ranks the relevant retrieved cases based on their local similarity. After that, the prototype by itself proposes a solution.

The proposed solution can be derived directly from a retrieved case that matches exactly or partially to the problem of the new case. Partially 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 may need editing (changing) based on different conditions. As a result the user of the system should have made an adaptation on the proposed solution having differences between the proposed case and the new case. In addition to adaptation, case contradictions are revised if there are situations where previous visitor's 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 can't provide recommendation to new cases. So during this time, this new case or problem of visitor 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.

In order to develop recommender system, the researcher collects essential knowledge from relevant documents, tourists and domain experts. In designing tourist attraction area recommender system, the researcher collects relevant attributes and cases from NTO as well as MoCT and the required knowledge was represented. Therefore, Building of case based recommender system was started by collecting previously solved cases (i.e. previous visitor's cases) from NTO & MoCT consisting of recorded data of visitors who are successful or satisfied in their recreational program. Since previously solved cases contains missing values and unnecessary information for this research, it need further processing in order to avoid such a problem and remove unnecessary attributes for tourist attraction area selection process. After processing of cases and selecting the most significant attributes, assigning weight and important parameters for each attribute was the next task which was performed. For the

selection of important attributes that influence the recommendation of best tourist attraction area and visiting time, the researcher used data mining attribute selection algorithm called attribute selection algorithm. The reason for using attribute selection mechanism is since all attributes are not equally important to recommend tourist attraction areas and suitable visiting time to new visitors.

Once the case based recommender system is developed, users/tourists can use the system easily to choose their attraction areas based on the recommendation given by the system in order to retrieve the best cases that can match with their query. When users/visitors enter their query/case description through the user interface window, the system searches the best matching cases from the case base and retains the possible solution. If there is exact matching between the query and previous cases in the case base, the system recommends the most matched attraction area and visiting time for visitors. If 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 visitors.

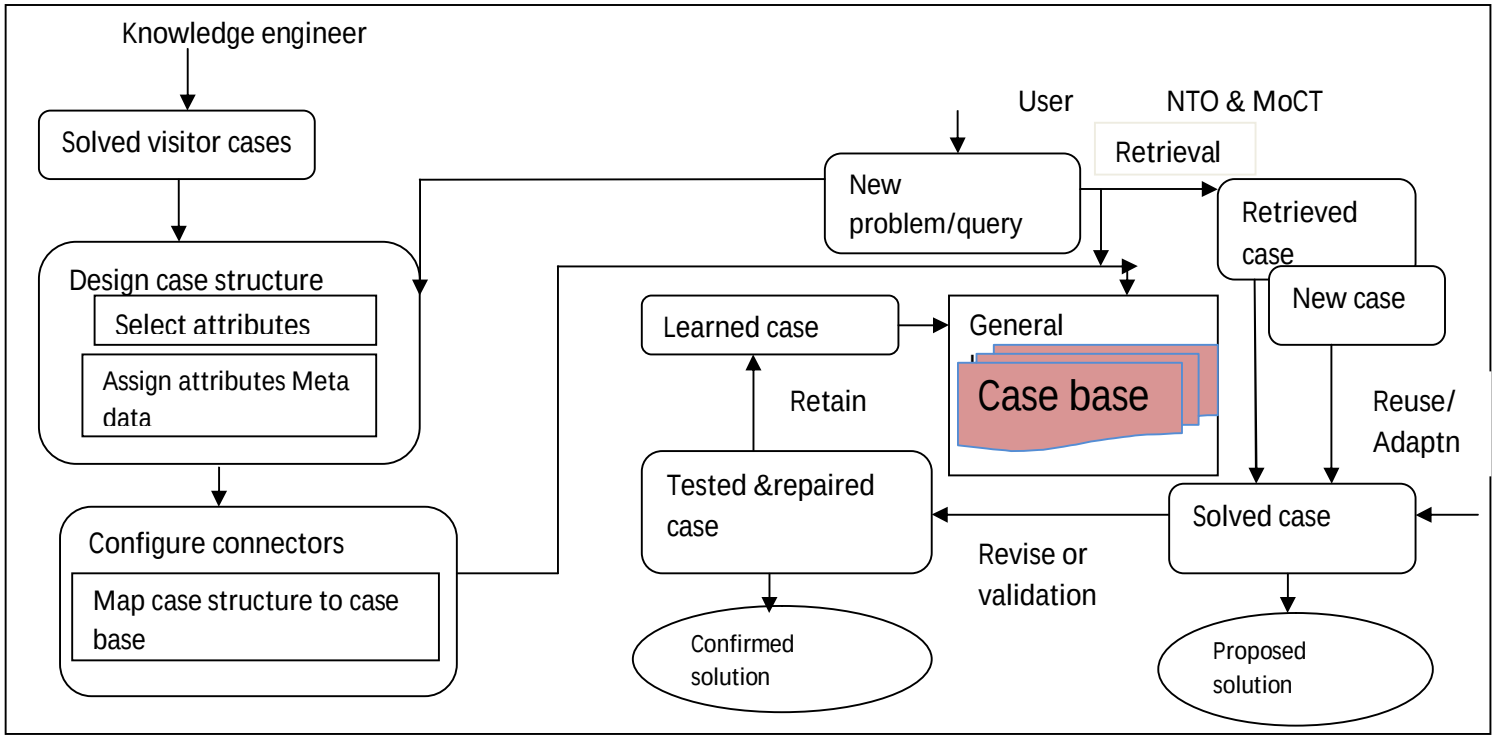


Figure 5.1 the architecture of CBR in tourist attraction area selection process

As it is shown in the architecture, when new cases which is not solved previously or not stored in the case base comes to the system, the system has a capability of learning from this new case. Then this new case can stored in the case library for future use and the case base updates in an incremental manner. After that, the number of cases which was stored in the case library increases due to the learning ability of the system.

5.3. Case Based Reasoning System for Tourist Attraction Area Selection

For the development of case based recommender system, the researcher used JCOLIBRI case based reasoning framework. JCOLIBRI has been constructed as core modules to offer the basic functionality for developing case base reasoning as well as case based recommender system.

As described by García, et al (2008), several steps are incorporated in the development of CBR application. Some of these steps are: collecting cases and background knowledge, modeling a suitable case representation, defining an accurate similarity measure, implementing retrieval functionality, and implementing user interfaces. For the purpose of this study, the researcher uses the main feature of jCOLIBRI to deliver the actual prototype. Implementing a CBR application from scratch remains a time consuming software engineering process and requires a lot of specific experience beyond pure programming skills (Juan A. et al, 2002).

Developing a new case base recommender system is made by writing few Java classes that extend classes of the framework and configure some XML files. To start the JCOLIBRI graphical user interface (GUI) application tool, launch the main window by clicking on **JColibriGUI.bat** file and it becomes ready to use as shown below in figure 5.2. The GUI of jCOLIBRI helps someone to create new CBR application with predefined task and methods. These predefined tasks and methods are represented in XML files that describe the tasks supported by the framework (tasks.xml) along with the methods for solving these tasks (methods.xml).

Figure 5.2: The Main Window of jCOLIBRI

The first and the primary task in the development of any CBR system using jCOLIBRI programming tools is instantiating the tool by double click on the tool itself and it will be ready for usage. After running jCOLIBRI GUI as shown in the above figure, the next task was **creating new CBR applications.** To develop a new CBR application, click on the CBR toolbar and then select "new CBR system". After that, a box which requires entering new application name will be displayed as shown fig 5.3 bellow.

5.4. Creating new CBR Application

New application can be develop step by step through the GUI of jCOLIBRI. To do this from the main windows of jCOLIBRI, i.e fig 5.2 click on CBR toolbar, then select new CBR system and give the name of new application (as Tourist Attraction Area recommender system) and press ok. After that, a new window will appear as shown in the figure bellow.

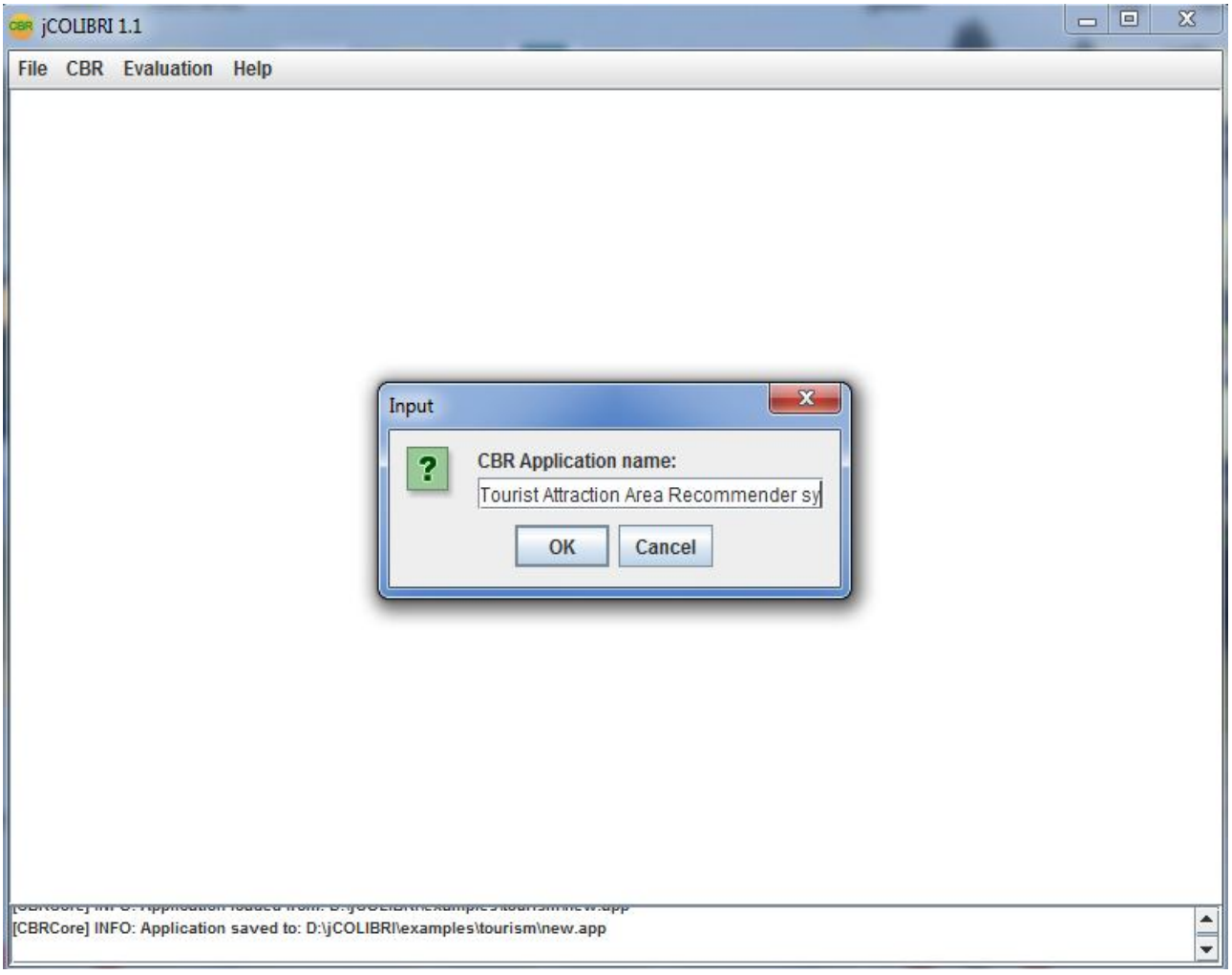


Figure 5.3 a screen shoot for CBR application start screen

After creating a new CBR application in such a way, a window will popup in which it asks to select one extension out of five as it shown in the fig. bellow. As noted by Mohammed H. (2006), the five extensions are: Core Extension, Case Retrieval Nets Extension, Description Logic Extension, Textual Extension and User Components Extension.

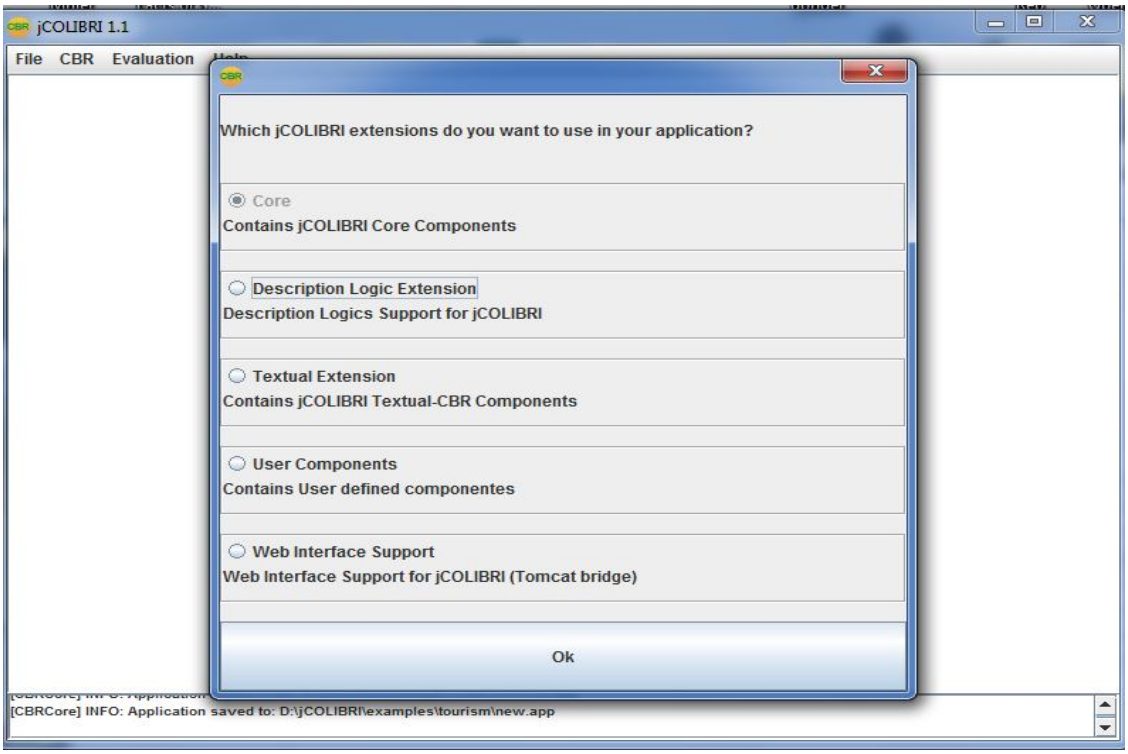


Figure 5.4 jCOLIBRI Extensions

Of these five extensions, for this study, the researcher selects core extension types because it contains all basics components of jCOLIBRI which is convenient for the development of case based recommender system. The 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, a user wants to define his/her own components, then, user has to select user component extension. After selecting core extension and pressing ok button, the basic CBR system window which comprises preCycle, CBR Cycle, and post cycle will be displayed as shown in fig.5.5 bellow.

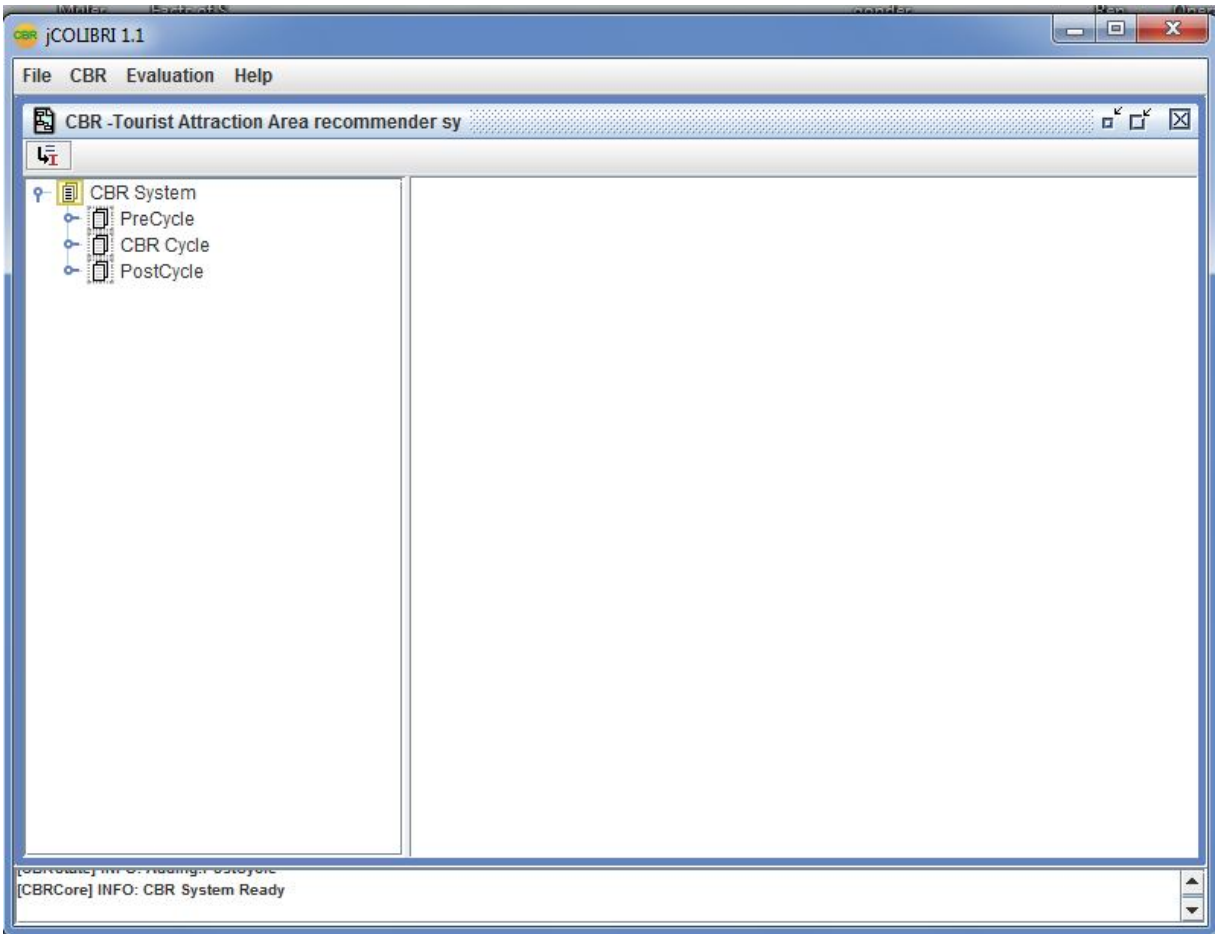


Figure 5.5 CBR Applications

Having created all the necessary CBR applications, in this study, tasks in the development of case based recommender system for the selection of tourist attraction areas can be partitioned in to different subsections like: building the case base, case representation, description of attributes, managing connectors and finally managing tasks and methods. The detail of each division is described here under:

5.4.1. Case Base Building

As described by Kriegsman, (1993), there are four basic steps in building any case-based reasoning system. First, data must be collected and represented. Second, the data must be examined, and significant features should be identified. Next, the data must be indexed for

efficient retrieval. Finally, the feature set and indexing scheme must be tested, evaluated, and revised as necessary to improve the system's accuracy.

Collecting the data "Good" case data, containing a representative and well-distributed set of historical cases, are the foundation of a good CBR system. Ideally, every area of the domain should be represented in the case base by a number of cases proportional to the complexity of that area. If the domain is viewed as a problem space, the cases should be evenly distributed and as dense as possible. Therefore, the researcher collects previous visitor cases from NTO to build the case base and represent it using suitable case representation method. The acquired cases are used to build tourist attraction area selection CBR system that can offer decision support to tourism experts, visitors and other entrepreneur professionals. 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 represented as a plaintext feature value representation comprising of N columns representing case attributes (A1, A2, A3... AN) and each M rows representing individual cases C ({C1, C2, C3, ...,CM}). Each attribute has a sequence of possible k values associated to each column attribute A={V1, V2, V3, ..., Vk}. 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 (selam, et al, 2005). The case base contains a set of cases that represents knowledge about tourist attraction area selection. The researcher tried to collect all the cases by selecting satisfied or successful visitors from NTO and MoCT due to right selection of attraction area.

Extracting features from the data: The next step was to establish a representation schema for the cases. The schema determines the level of detail the system can inspect about each case, and the level of granularity at which the cases can be indexed. This directly influences system accuracy and precision.

Indexing the data The third step in building a case-based system is to select an index and retrieval scheme that suits the application requirements and the features available in the data.

Indexing refers to assigning index to the case for retrieval by comparing the existing case and the query given by the user.

Testing and refining the indexing schema: as cases are added to the library and cluster tree are built, it becomes important to measure the accuracy of the system to see if it needs further tuning.

5.4.2. Case Representation

A case in case-based reasoning is a contextualized piece of experience, which can be represented in various forms and that teaches a lesson fundamental to archiving the goal of the reasoner. The issue of case representation involves deciding the type and the structure of the domain knowledge within the cases. Case representation is one of the main components in case based reasoning system, because it is the process of representing the case in the way that, the programming language (jCOLIBRI) can easily understands or interpret it. Case representation in case-based reasoning (CBR) makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. A large variety of representation formalisms have been proposed. However, three major types of case representation have arisen: feature vector (or propositional) cases, structured (or relational) cases, and textual (or semi-structured) cases. (Bergmann et al, 2005). Of these methods of representations, the researcher used feature vector case representation mechanisms because it represent a case as attribute-value pairs, similar to the propositional representations used in Machine Learning (ML), that support k-nearest neighbor matching and instance-based learning (Bergmann et al, 2005).

Creating a case structure is an important phase of any CBR system development because it helps to define the features available in the cases and used to measure the similarity between existing cases and the new case (query). So, the whole application of this study is to retrieve similar cases from the case base that can show future reasoning, problem solving, transforming a solution retrieved into a solution appropriate to the current problems (i.e. to retrieve similar cases to the query from the case base that guide visitors), and making a

recommendation in tourist attraction area and appropriate visiting time selection process. Case base were structured to make the retrieval process efficient. Case retrieval is a process that a retrieval algorithm retrieves the most similar cases to the current problem. This is done through case indexing process in the JCOLIBRI programming tool. An index is a computational data structure that can be stored in memory and searched quickly. Case indexing involves assigning indexes to cases to facilitate their retrieval by comparing the existing case and the query given by the user (Cunningham and Zenobi, 2001).

5.4.3. Description of Case Attributes

Typically, there are three major parts of a case; these are: **Problem description**: the state of the world while the case is happening and what problem needed solving at the time. **Solution**: the stated or derived solution to the problem. **Outcome:** the resulting state of the world after the case occurred. Description and solution are set of simple or compound attributes, permitting us to build a hierarchical case structure. In jCOLIBRI, the case structure can be defined by using manage case structure window (Cunningham and Zenobi, 2001). Description of attributes is a means of describing an attribute or manages the case structure which is used for giving recommendation for tourist attraction area selection. Attribute description can be done by using add simple button or compound description attributes in description case structure and we can set the properties of attributes or metadata of attributes for each description attributes. Metadata of attributes includes weight of attribute, data type of attribute and similarity function. Here, before creating CBR application we need to configure the case structure. To do this click on the toolbar menu button of CBR and select the option mange case structures. Then a window will appear for configuration of case structure. In fig 4.6 bellow, left side's options are case structure description, solution and result. Remove button will remove the cases. When we click on case like (Recommended attraction area) then its properties will appear on right side. Name, type, weight and similarity of case can be changed in properties of the case. Apply changes button will apply these changes. This will impose on case. Local similarity is used for computing the similarity of attributes. After defining the structure of cases, case structure is saved automatically in xml file.

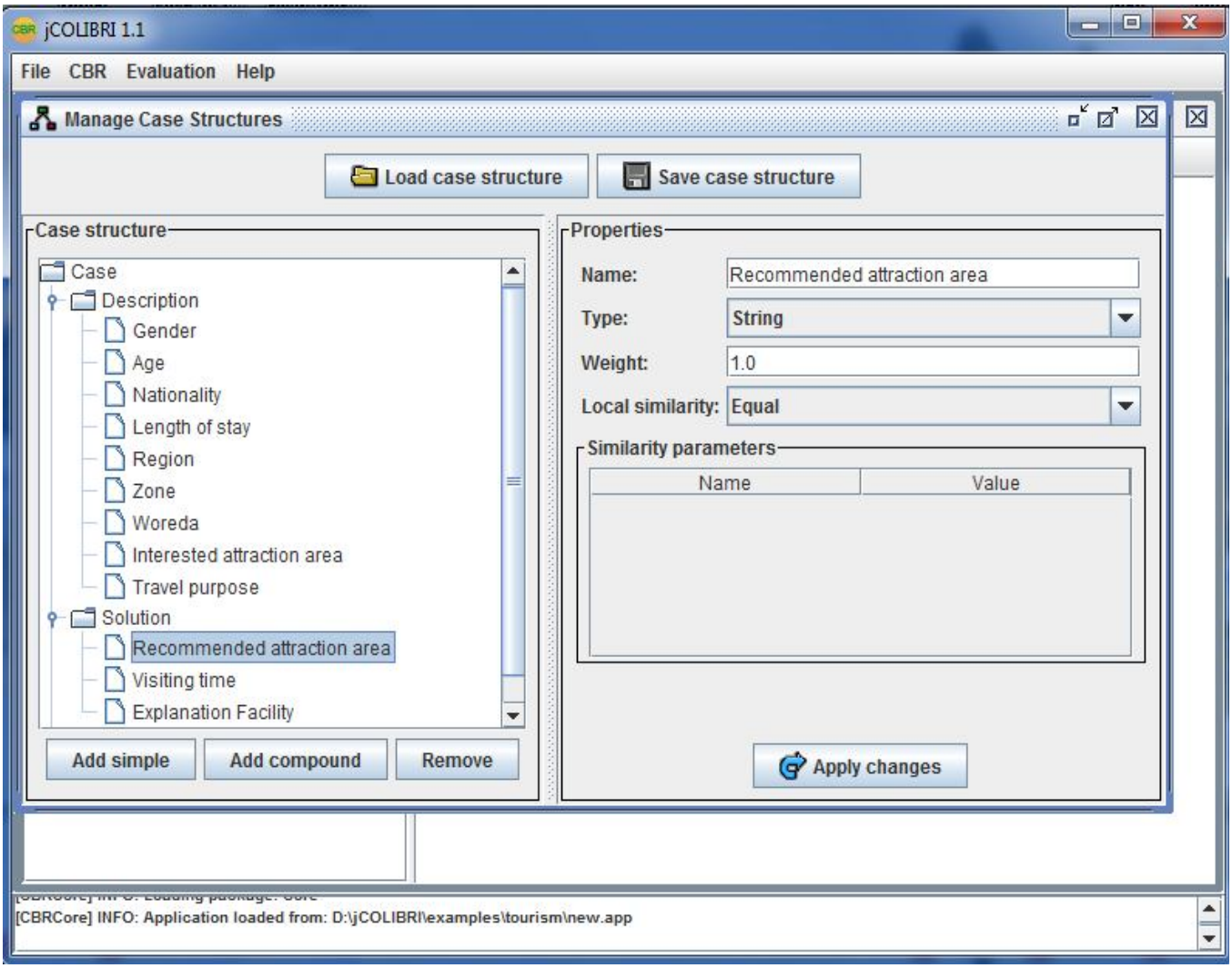


Figure 5.6 Case structure defining and similarity

As shown in fig 5.6 above the case structure of this research contains, nine description attributes which consists of descriptions of the problem needed. And there are three solutions attribute to make sound decision by the system. Solution attribute is assigned to the new case (visitor) after they supply the value of all description attributes and measuring the similarity between the existing case attribute value and new case attribute value. For this research the solution attributes include recommended attraction area, recommended visiting time, and explanation facility about the recommended attraction area. The following table shows the description of selected cases with their value and Weight.

Table 5.1demonstrate the description of case attributes and solution attribute on the subject of name, data type, weights, local and global similarity.

Significant Attributes			
Attribute Name	***Data type***	***Weight***	***Local similarity***
Gender	String	1.0	Equal
Age	String	1.0	Max string
Nationality	String	1.0	Equal
Length of stay	Integer	1.0	Interval
Region	String	0.9	Max string
Zone	String	0.9	Max string
Woreda	String	1.0	Max string
Interested attraction area	String	1.0	Max string
Travel purpose	String	0.95	Max string
Solution			
Recommended Attraction area	String	1.0	Equal
Recommended Visiting time	String	1.0	Equal
Explanation Facility	String	1.0	Equal

Table 5.1 description and weight of the selected attributes

The above table shows the general description of attributes consisting of attribute name, data type, weight and local similarity. As shown in the table 5.1 the most significant and discriminating attributes to the problem domain has the highest weight value (that is 1.0). These attributes are the most relevant to visitors to select convenient attraction area and visiting time that can fit with their personal and socio-economical characteristics based on their interest. Next to these, attributes like region and zone has a weight value of 0.9 each and the attribute travel purpose has a weight value of 0.95. The assignments of weights to each attribute indicates that attributes having high weight is the most relevant to the user (visitors) in the selection process of tourist attraction area and appropriate visiting time. The weight

value of each attribute has been assigned by using information gain attribute selection algorithm and by the help of domain experts. The local similarity of most description attributes is maximum string. This is due to the similarity between query and cases can be calculated with maximum string length. Few attributes such as gender and nationality, have equal similarity weight since local similarity needs exact match of existing cases and new case (query).

Having identified the most relevant attribute of the case, the next task is definition of appropriate similarity measure in JCOLIBRI. The programming tool (JCOLIBRI) follows both local and global similarity measures. The various local and Global similarities used in this research are discussed below:

A.Local Similarity: this kind of similarity measure divides the similarity definition into a set of local similarity of each attribute. There are three types of local similarity measurement :

I. ***Equal*:** If we select equal local similarity for each attribute, then our input and value of case base must be exact match. If the value between attribute are exactly match, the system gives (assign) a solution or recommended attraction area and visiting time to new cases. Otherwise matches are a failure and have no any solution or recommendation to the new query.

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

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

IV. ***Threshold:*** You have to set value in threshold as interval. It also compares values keeping in mind threshold.

B.Global Similarity: it is associated with compound attributes and used to get similarity of collected attributes in unique similarity value. Global similarity calculates the final similarity measure. The type of global similarity used in this research is average similarity. Average similarity is a type of global similarity that considers the average of all attribute of local similarity

values. Here, the local similarity of all case attributes which have string data type have either equal or Max String similarity value where as the Global similarity of all case attributes which have any data type have average similarity value.

5.5. Managing Connectors

The CaseComponents which is defined in the previous section are used to fill the attributes of the CBRCase objects when loading the cases. The CBR systems must access the stored cases in an efficient way from the case base, a problem that becomes more relevant as the size of the Case Base grows. jCOLIBRI manages that problem in to two ways these are: Persistency mechanism and In-memory organization.

Persistency mechanism: Persistency is built around connectors which represents first layer of jCOLIBRI. Connectors are objects which know how to retrieve cases from case base and how cases can be returned to CBR system in efficient way. jCOLIBRI implements various types of connectors: such as File system connector's retrieve cases from XML files, JDBC connectors make possible for jCOLIBRI to use with DBs in the market and RACER connectors which allows the designer to use RACER. RACER is responsible for maintaining and giving access to cases.

In-memory organization: An interface which assumes that whole Case Base can be read into memory for the CBR to work with it. It is not feasible for big size. That is reason, for which jCOLIBRI connectors implemented a new interface who allow retrieving cases satisfy a SQL query. Designer can decide when and what part of cases loaded into memory. Second layer of Case Base is Data structure which will organize cases after they loads into memory.

Therefore, managing connector performs the task of configuring the connector that is going to load the case base. JCOLIBRI supports both SQL database and plain text file to store its cases base. It also splits the problem of case base management in two separate although related concerns: persistency mechanisms through connectors and in-memory organization.

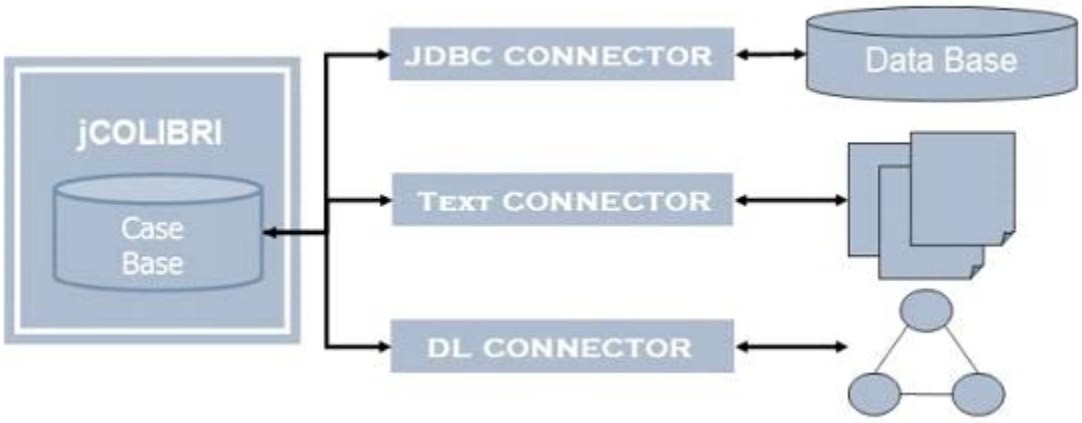


Figure 5.7 JCOLIBRI connector schema (source: García, et al, 2008).

As it is depicted in the figure above, Persistence is built around connectors. A connector represents the first layer of jCOLIBRI on top of the physical storage. Thus, Connectors provide an abstraction mechanism that allows users to load cases from different storage sources in a transparent way. The use of connectors give jCOLIBRI flexibility against the physical storage so the system designer can choose the most appropriate one for the system at hand. As shown in figure 5.7, jCOLIBRI includes connectors that work with plain text files, relational databases and Description Logics systems. For this study, the researcher used plaintext type of connector because visitor's cases are stored in plaintext file format as shown fig 4.8. Plaintext file case base connector is used for persistence of cases. In this connector, the researcher has to specify the path of case structure and also path of text file. All the attributes of a case should be mapped. 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. As a delimiter of this connector, comma (,) was used to separate value of each attribute in the case. This is connector's responsibility to retrieve data from case base and return it back to GUI. Like that of case structure, connector is also saved in xml format.

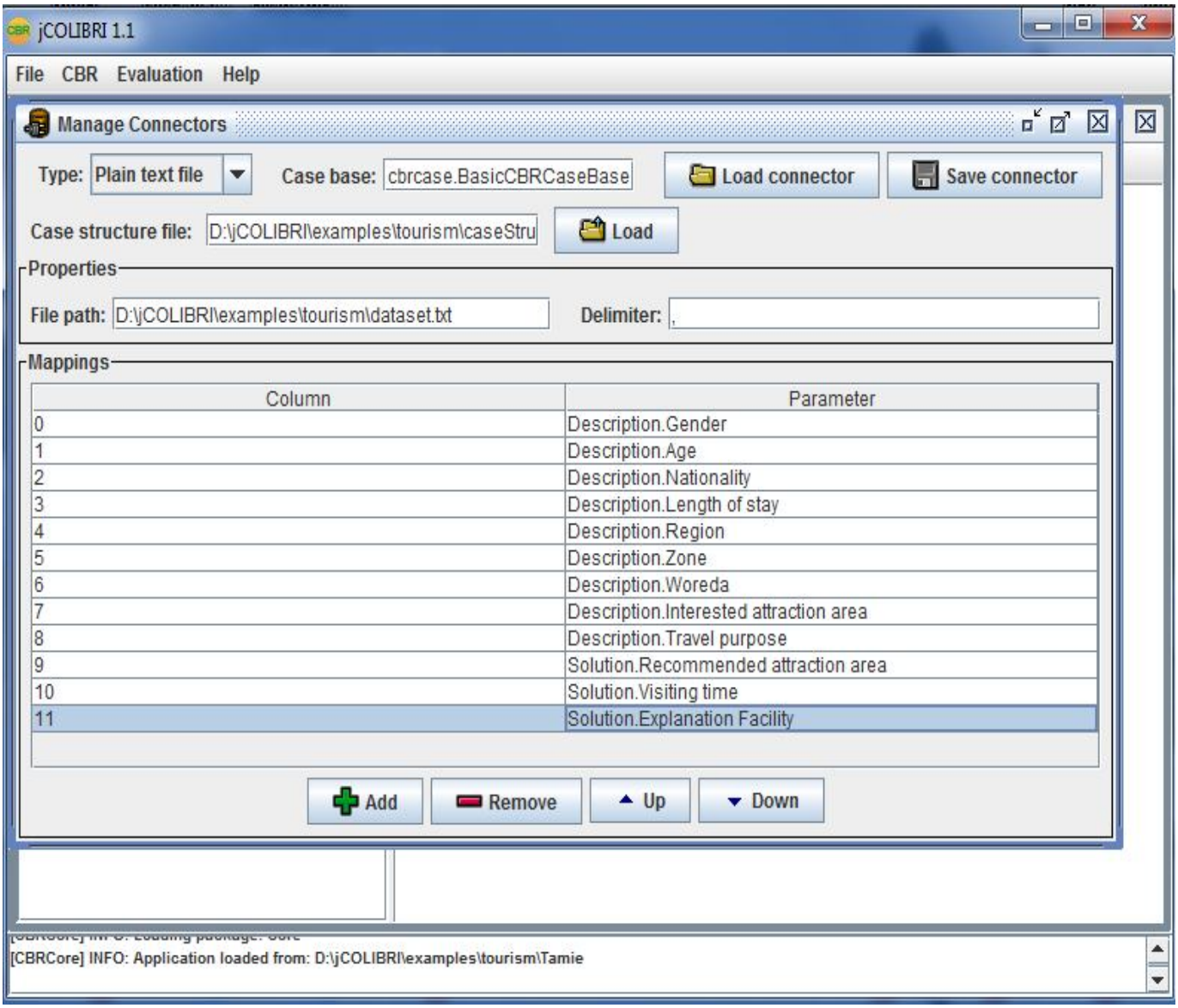


Figure 5.8 managing connectors

5.5.1. Managing Tasks and Methods

The programming tool JCOLIBRI is organized into packages. These packages can perform and execute tasks and methods decomposition process. For the development of case base recommender system prototype, the researcher used core package task. Although, various core packages are available in JCOLIBRI, the main components of Core packages which are used for the development of this prototype recommender system are PreCycle, main CBR

cycle and PostCycle. The detail of each tasks and methods can be discussed separately as follows.

5.5.1.1. Managing Tasks

After configuring the connector and case structure, the next step is selecting tasks and methods of the application. jCOLIBRI has two types of task packages, namely, Core packages and User defined package tasks. For the implementation of this prototype recommender system, 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.

A core package also have predefined tasks and methods that are 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, because defined tasks and methods by user itself for every system is time taking and complex. The component of core packages is the final and important step for creating a new application where the CBR application is configured. The left side of Figure 5.9 bellows shows PreCycle, CBR Cycle and PostCycle.

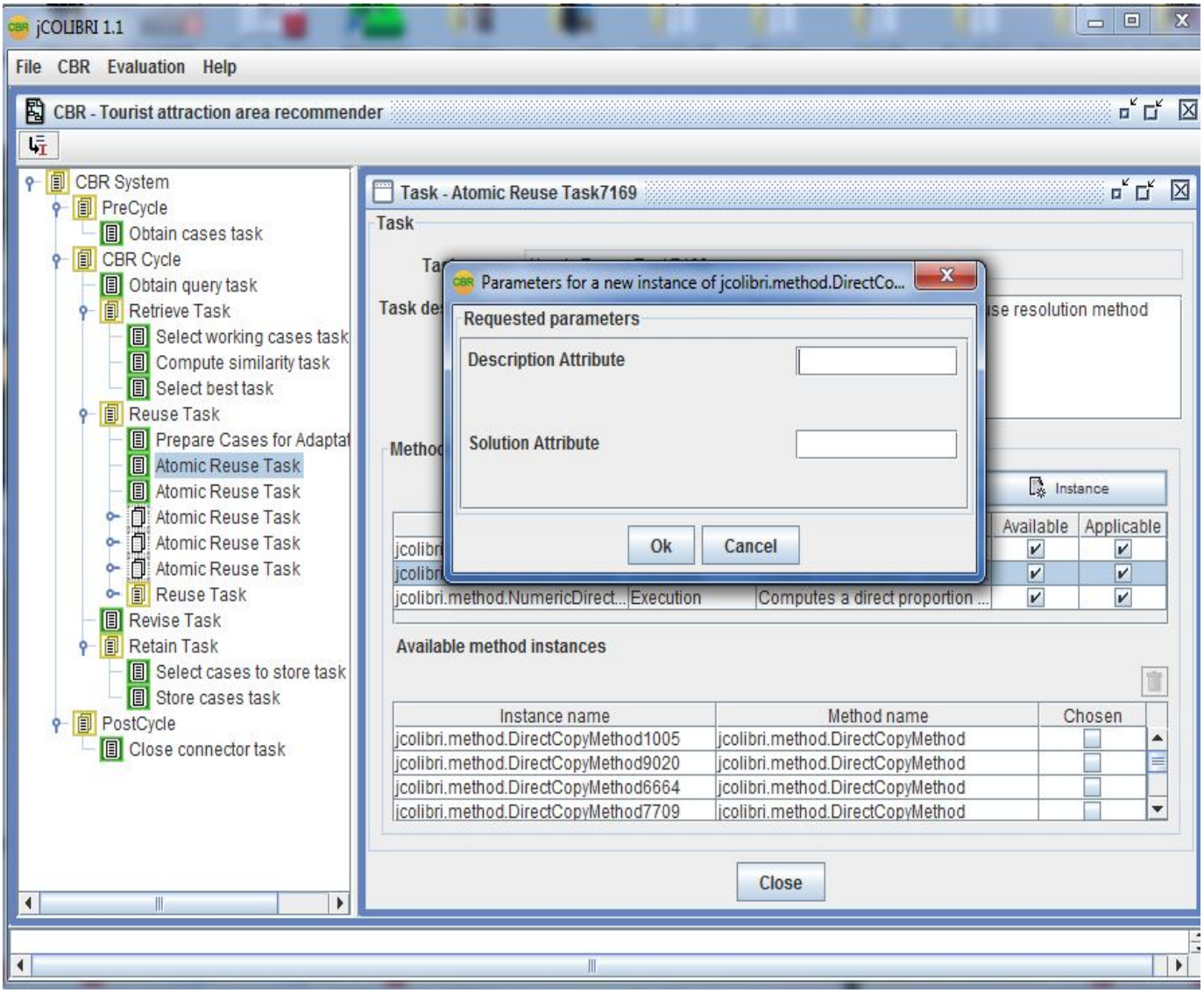


Figure 5.9 configure the CBR application

As it shown in the figure 5.9 above the main tasks and activates on each CBR systems or components of core packages can be expressed as follows:

PreCycle: This part of the task retrieves data or cases from case base before execution of the main cycles. Its task is to get all the cases in case base. Hence, the component of the core packages the researcher start with "PreCycle" in order to loads the cases from data sources (case base). In preCycle tasks are solved once before the main cycle, like computing the index structure or processing texts in textual CBR. Therefore to load the cases from the case base it

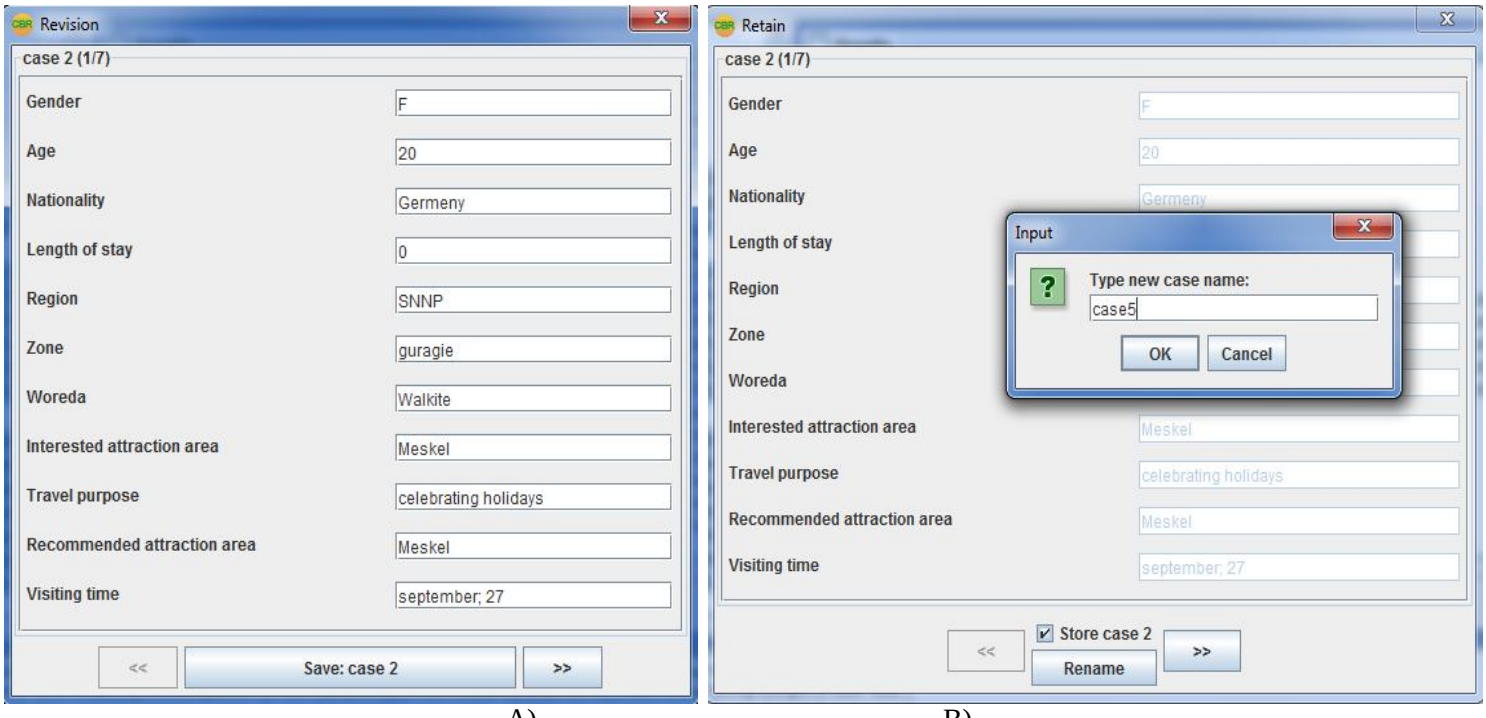
is necessary to define the path of the connector on subtask of Precycle called "obtain cases task" and make "instance" to instance the tasks and methods. In Precycle task there is only one subtask called "obtaining case task". Obtain case task is used retrieve (load) cases from visitor's case base before the execution of the main CBR cycle.

Main CBR cycle: this cycle retrieves the most similar tasks and describes the typical cycle task at the highest level and obtains the query. It 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 of visitor's case attributes that are available after he path is assigned. In addition to obtaining query task, there are other significant tasks under the main CBR cycle. These are retrieve tasks, reuse tasks, revise task and retain tasks.

- **Retrieve task:** is used to retrieve case(s) from the stored case base. Retrieve tasks is also decomposed in to different subtasks. These 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(s) after computing the similarity of stored cases against the new case. It means that the number of best matched case(s) is shown to the user depending on the method used and the threshold.
- **Reuse/Adaptation tasks:** it enables 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. But there are situations where previous visitor's cases are not similar with the new case or the problem of visitor, so during this time, this new case or problem of visitor can be stored in the case base and will be reused by other visitors for the

next time. The system can learn at every entry of new case and new users adopt this knowledge for tourist attraction area and convenient visiting time selection process.

- **Revise task:** it is the evaluation and correction stage about the recommended solution in reuse phase. As shown in fig 5.10(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** Its main aim is 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 fig 5.10 (b). Select cases to store task give authentication to the user for storing case. The store cases task enables to store case(s) 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.



A) B)

Figure 5.10 Revision and Retain tasks

PostCycle: is the last task in managing tasks in jCOLIBRI. PostCycle 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.

5.5.1.2. Managing Methods

The method 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.

The most prominent methods which are used to solve tasks for this CBR application are described here under:

- **LoadCaseBaseMethod**: This method returns the whole available cases from the case base to designer. It also uses connector to retrieve case base.
- **ConfigureQueryMethod:** 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.
- **SelectAllMethod**: This method allows displaying all the available cases from the case base to the result window. It also Selects working cases from case base and store them into current context.
- **NumericSumComputationalMethod:** it Computes similarity tasks between cases and query.
- **NumericProportionMethod**: it is the sub method of reuse task which involves in computing numeric proportion between the description attributes and solution attributes.
- **ManualRevisonMethod**: this method enables users to modify cases in the query window as they wish.
- **SelectSomeMethod**: select best of found cases. It returns most similar value of the top best selected case. This method resolves to select best task by choosing the "n" number of nearest exact similar cases from the returned cases. The "n" indicates since there is more than one similar (relevant) case to the new case (query), the system retrieve "n" number of similar cases in ranking order of from highest similarity to lowest similarity. These method requests

the users to enter the value of each query as input. Then the system measures the similarity between the new queries input value and the existing case value.

Generally, Tasks in JCOLIBRI programming tool 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 study, only few of them are selected and discussed which are appropriate for recommender system.

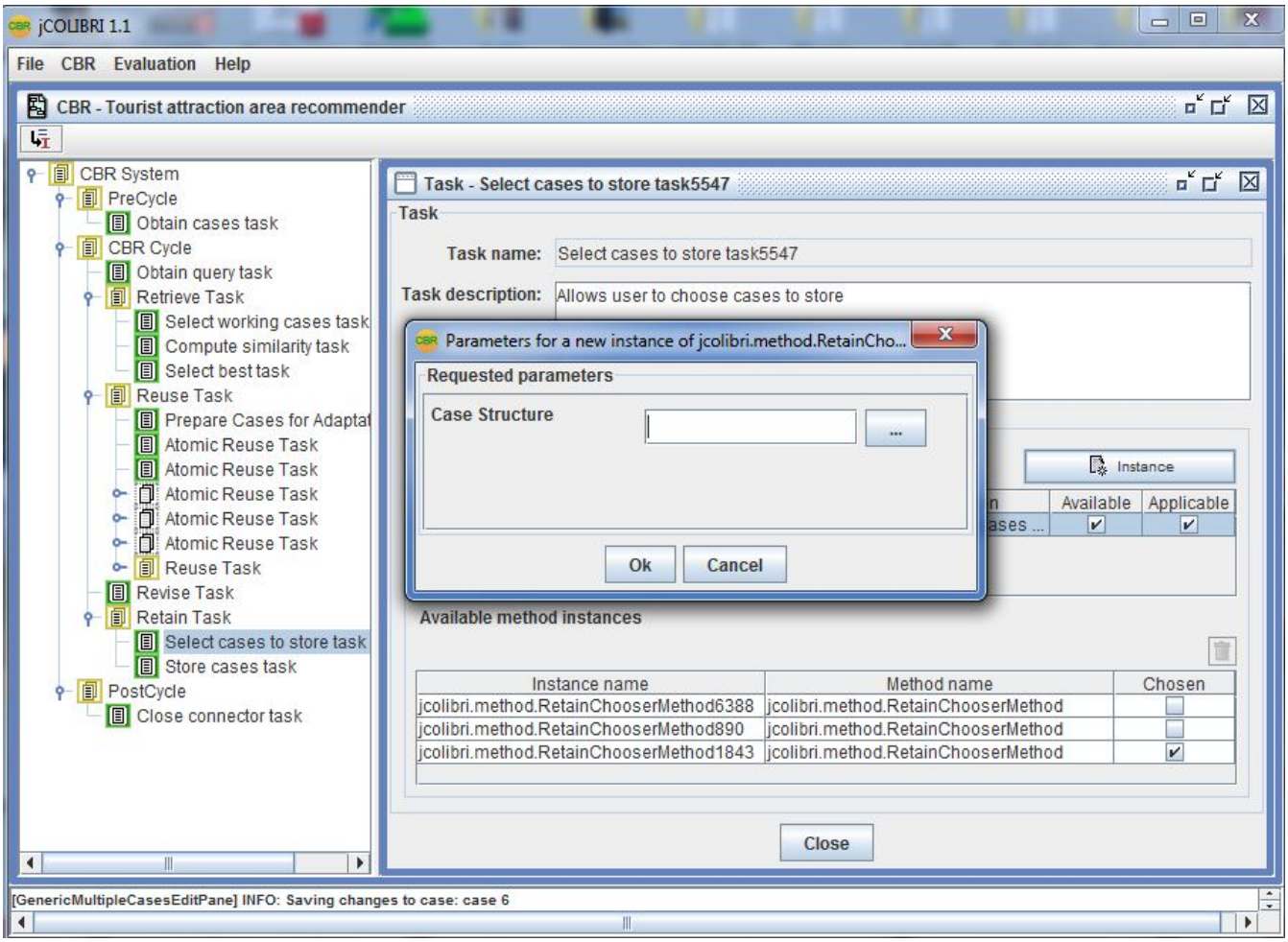


Figure 5.11 managing tasks and methods

As shown above in Figure 5.11 the main window of cycle of JCOLIBRI tasks and methods can be stated in such manner and the left side shows the tasks and subtask where as the right side shows the methods. The figure depicts, 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 circumstances. These inputs are parameters for new instances.

5.5.2. Similarity of cases, Matching and Ranking

The key objective of case base recommender system is to retrieve the best similar cases to the query from case base and selecting the nearest similar case by using some similarity assessment of heuristic functions. The similarity function involves computing the similarity between the stored cases in the case base and the new cases (query), and selects nearest similar cases to the query. Hence, the programming tool jCOLIBRI uses the nearest neighbor algorithm as a case retrieval technique. Nearest neighbor algorithm is used to measure the similarity between the stored (existing) cases and the new cases (queries), and return the search results within their ranked order. For each and every attribute in the query and case, local similarity function measures the similarity between each and every simple attribute values in the case base with new case queries. Based on the matching weighted sum features from those simple attributes, the similarity score between the queries and stored cases for each simple attribute is assigned.

At the end, the average score (global similarity) of each attribute between the existing case and the query are computed and the result is assigned to the object (the similarity between the stored case and the query). And then the maximum degree of similarity among the retrieved cases is displayed according to their ranked order. Figure 5.12 bellow demonstrates how similarity of cases are computed the case base and the query.

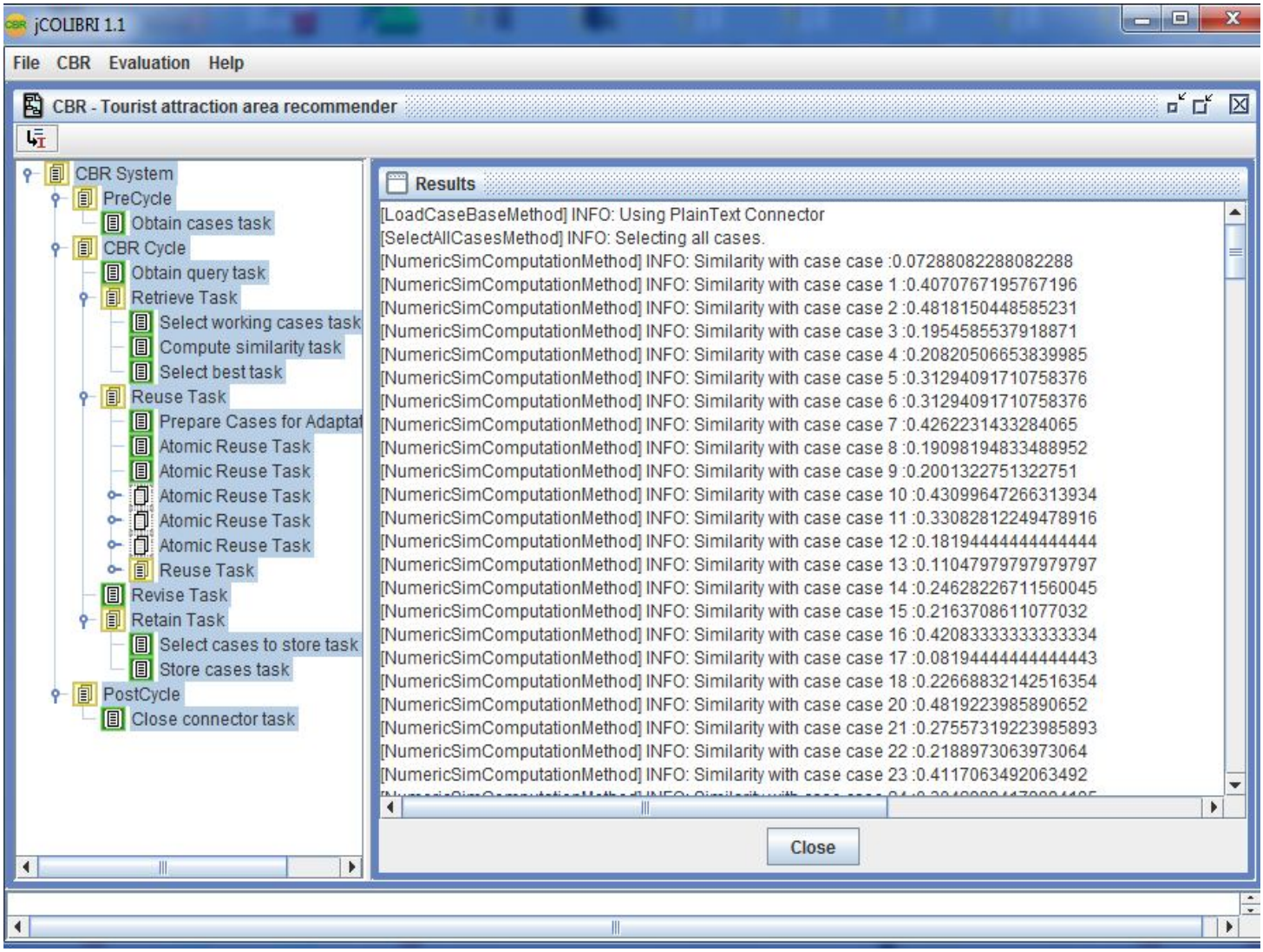


Figure 5.12 case similarities between case base and query.

5.5.3. Deploy the Case Based Recommender system

After defining and configuring all the necessary steps required in designing case based recommender system using the programming tool JCOLIBRI, new case (query) entry application for new tourists is the next step as shown in figure 5.13 bellow.

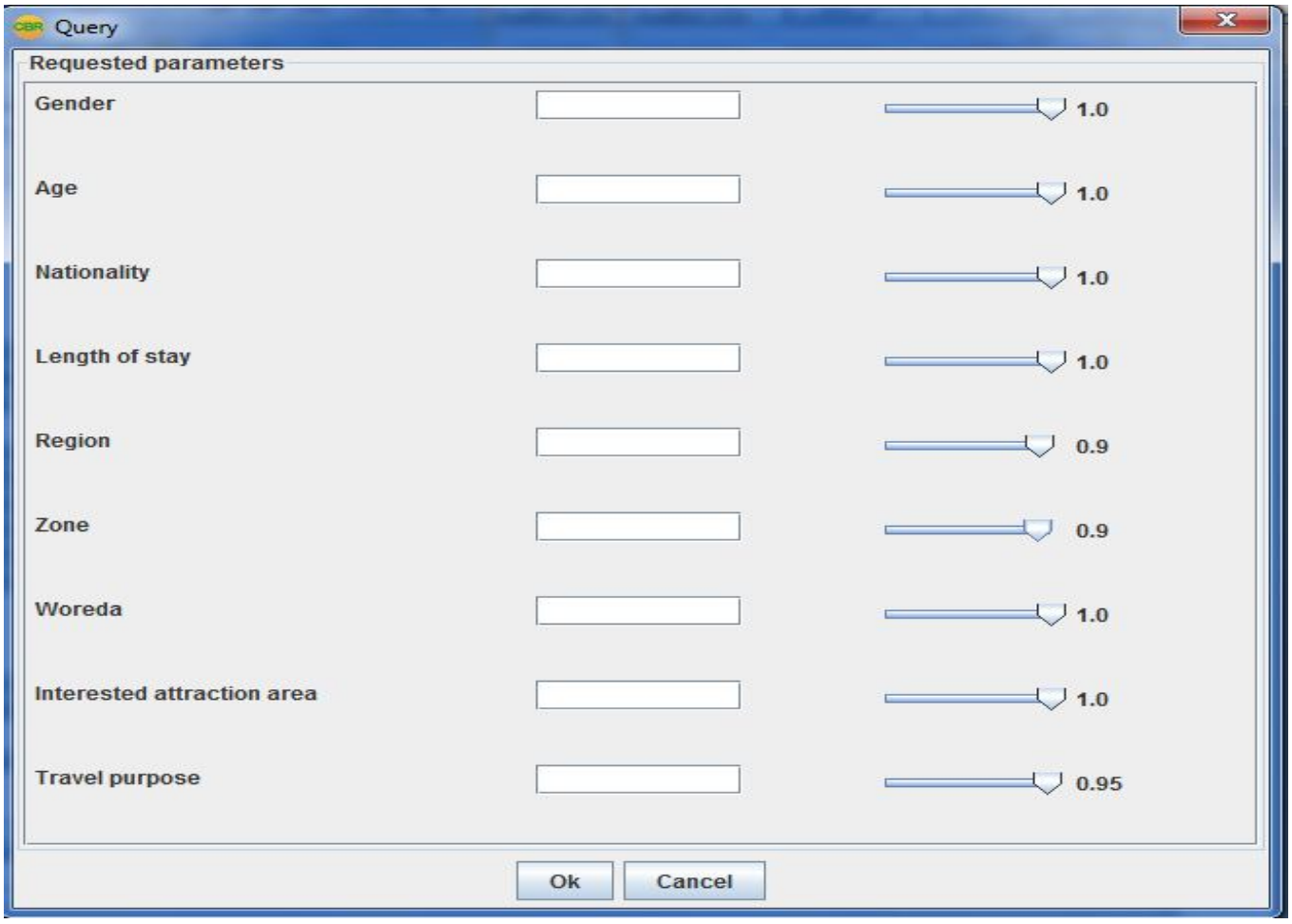


Figure 5.13 windows for case entry in to the case base

In the above query window visitors are expected to enter the query to each requested parameters or attributes in the space provided. After entering the query, at the bottom of the screen they will see the results of similar previous visitor cases and the recommended attraction areas, recommended visiting time, and explanation facility about the attraction areas on the execution log. For example in "Nationality" box visitors are required the query of their ethnic group as Ethiopian, Germeny, Itally, Spain etc.

In “length of stay” box tourists are also required to supply the duration of time in which they could be able to stay i.e in the country like one week, three months, 2 nights etc. In “interested attraction area” box visitors are required to input their interest they want to visit. The location of attraction area is entered by tourists/visitors such as regions of attraction area location, zones of its location and weredas of attraction areas location in different box. At the end visitors are also expected to feed their travel purpose (the reason why they come to Ethiopia and they can supply as vacation, business, visiting relatives, conference etc).

5.5.4. Explanation Facility

One of the more interesting features of knowledge based systems is their ability to explain themselves. The explanation facility in this study is used to give explanation about the recommended attraction area after decision or recommendation is made by the system. Once the system reaches its final decision on the recommendation of attraction area and appropriate visiting time, the user may not have brief information about the recommended attraction area. In this case the system provides explanation facility about the recommended attraction area. Then the system gives more descriptions about the attraction areas such as the definitions, location, type of accommodations while visiting etc.

CHAPTER SIX

6. TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE

6.1. Introduction

Testing and performance evaluation of the prototype case based recommender system is the final step that helps the researcher to measure whether the system achieves the proposed objectives or not. This chapter presents performance evaluation of the prototype system. For the performance evaluation of this research has conducted user acceptance testing of the prototype, case similarity testing, and retrieval performance evaluation using recall and precision.

6.2. User Acceptance Testing

User acceptance testing is performed in real circumstances at National Tour Operation specifically with Tourism Experts and visitors. During testing the user acceptance, the applicability of the prototype is evaluated by potential users of the system. Potential users of the system are tourism experts and visitors. Case based reasoning system user acceptance evaluation method allows users (domain expert and visitors) to directly interact with the system and evaluate the performance of the case based system from the users' point of view. User acceptance testing helps to ensure the performance of the prototype by assessing the feedback acquired from the domain expert and visitors towards the developed prototype system.

In order to evaluate the user acceptance of the developed prototype system, the researcher used questionnaire adapted from Getachew (2012), Biazen (2013) and Yibeltal (2013). To achieve the ambition of user acceptance evaluation of the prototype system, fourteen visitors and eight domain experts from NTO who were participating in different tourism sectors in the country were purposely selected. Throughout the development of case based

recommender system, these domain experts were actively participated in different phases of the study, including knowledge acquisition and prototype development. Before starting the evaluation process of the system using the questionnaire, the researcher first gave explanation about the system to domain experts and visitors in NTO. The explanation given for experts and visitors assists them to avoid the variation of awareness among them about the prototype case based recommender system.

Following the explanation, domain experts and visitors/tourists were allowed to interact with the system by running number of cases having similar parameter with the facts incorporated in the case base. After the consultation of the system, to assess the user acceptance of the prototype case based recommender system, nine close-ended questions were distributed to domain experts and visitors. The first three questions are on the user interface design aspect which is basic for users interface satisfaction. These questions assessed whether the user interface of the system is easy to use, attractiveness and time efficiency of the system. The rest of the questions are used to evaluate the prototype's adequacy and clarity, relevancy of retrieved cases, relevance of the attributes used, clarity of the explanation facility, problem solving ability and significance of the prototype knowledge based system in tourist attraction area and visiting time recommendation system. All these nine closed ended questions answered as excellent, very good, good, fair and poor. For the ease of analyzing the performance of the system based on user's feedback, the researcher assigned numeric values to the five options as follows: excellent=5, very good= 4, good=3, fair=2, poor=1. The system evaluators gave the value for each closed ended questions.

The Table below depicts the feedbacks obtained from the domain experts (evaluators) on systems interaction as calculated based on the given scales.

No	Evaluation criteria	Performance value						
		1	2	3	4	5	Average	%
1	Easy to use of the recommender system				5	3	4.4	88
2	Is the system efficient in time				4	4	4.0	40
3	Is the user interface interactive			2	5	1	3.4	68
4	Adequacy and clarity of decision support			3	4	1	3.8	76
5	Relevancy of the retrieved case in the decision making			1	4	3	4.3	86
6	Fitness of the final solution to the new case			2	4	2	4.0	80
7	Relevancy of the attributes in representing visitors case			3	3	2	3.9	78
8	Does the explanation facility give brief description about the recommended attraction area			2	4	2	4.0	80
9	Rate the significance of the system in the domain area				2	6	4.8	96
		Total Average					4.0	80%

Table 6.1 performance evaluation by domain expert

As shown in table 6.1, 62.5% of the respondents rate the easiness of the recommender system as very good and the remaining 37.5% of the respondent's rate it as excellent. At the same time, efficiency of the system in terms of time is rated by domain experts as 50% very good

and 50 as excellent. In case of user interface interactivity of the prototype, 25% of the respondents rated as good, 62.5% rated as very good and the remaining 12.5 %rated as excellent. 37.5% of the respondent rate adequacy and clarity of the system as good and in the same way 50% and 12.5% of the respondents are rated as very good and excellent respectively.

Similarly, relevance of the retrieved cases in decision making to support users by recommending most similar cases rated as good by 12.5% of the respondents, very good by 50% of them and the remaining 37.5% of the respondents rate it as excellent. Fitness of the final solution to the new case is also rated by 25% of the respondents as good, 50% of them as very good and the remaining 25% of the respondents as excellent.

37.5% of the respondents rate Relevancy of the attributes in representing visitors' case as good, 37.5 % of them as very good and the remaining 25% of the respondents as excellent. In the case of explanation facility, 25% of the respondents rated as good, 50% of them rated as very good and the remaining 25% of the respondents rated excellent. Finally, 25% of the respondents rate the applicability of the prototype in their domain area as very good and the remaining 75% of the respondents rated as excellent.

No	Evaluation criteria	Performance value						
		1	2	3	4	5	Average	%
1	Easy to use of the recommender system			3	9	2	3.9	78
2	Is the system efficient in time				8	6	4.4	84
3	Is the user interface interactive				12	2	4.1	82
4	Adequacy and clarity of decision support			4	7	3	3.9	78
5	Relevancy of the retrieved case in the decision making			2	9	3	4.0	80
6	Fitness of the final solution to the new case			3	7	4	4.0	80
7	Relevancy of the attributes in representing visitors case			5	8	1	3.7	74
8	Does the explanation facility give brief description about the recommended attraction area				11	3	4.2	84
9	Rate the significance of the system in the domain area				1	13	4.9	98
		Total Average					4.1	82%

Table 6.2 performance evaluation by visitors

On the other hand, table 6.2 above shows the performance evaluation of the prototype by the visitors/tourists.

As depicted in figure 6.2 above, 21.428% of the respondents rated the easiness of the recommender system as good, 64.285% of them rated as very good and the remaining 14.285%

rated as excellent. In case of efficiency of the system in terms of time, 57.14% of the respondents were rated as very good, and the remaining 42.85% of them rated as excellent. Regarding to the interactivity of the prototype, 85.71% of the respondents rated as very good and the remaining 14.28% of them rated as excellent. In the case of adequacy and clarity of the system, 28.57% of the respondents rated as good, 50% of them rated as very good and the remaining 21.42% of the respondents rated as excellent.

In the same way, Relevancy of the retrieved case in the decision making process was rated as good by 14.28% of the respondents, very good by 64.28% of them and excellent by the remaining 21.42% of the respondents. Again 21.428% of the respondents rated fitness of the final solution to the new case as good, 50% them rated as very good and 28.57% of the respondents rated as excellent. The relevancy of the attributes in representing visitor's case is also rated by 35.71% of respondents as good, 57.14% of them rated as very good and the remaining 7.14% of the respondents rate it as excellent. While 78.57% of the respondent rate explanation facility gives brief description about the recommended attraction area as very good and the remaining 21.42% rate it as excellent. At the end, 92.85% of the respondents rate the applicability of the prototype in their domain area as excellent and the remaining 7.14% of the respondent's rate as very good.

6.3. Case Similarity Testing

To know how new cases are matched with the cases from the case base; an experiment was conducted by taking four experimental groups. The first group is made up of cases from the case base. The second group consists of cases which are made by modifying one of the attribute values of the case from the case base, while the third group is made up of cases which have two modified attribute values, and the fourth group consists of cases with three modified attributes values. Each test case is presented to the system individually to evaluate the performance of the similarity measures .Table 6.3 below shows the sample of queries that are used in this experiment with their values.

	Age	Gender	Length of stay	Nationality	Travel purpose	Interested attraction area	Region	Zone	Wereda
Query 1	39	M	Two weeks	USA	Vacation	Semein mountain	Amhara	n/Gonder	Debark
Query 2	22	F	One month	Itally	Conference	Lalibela	Amhara	n/wollo	Lalibela
Query 3	40	M	Three nights	Germeny	Transit	Axum obelisks	Tigray	Axum	Axum
Query 4	53	M	Three weeks	France	Visiting relatives	The walled city of Harar	Hrari	Harari	Harari
Query 5	19	F	Two nights	China	Vacation	Tia's carved standing stones	SNNP	Guraghe	Butajira

Table 6.3 Sample of queries that are used in this experiment with their values

Based on the above attributes, the next step is doing the experiments for the four groups. The number of cases used to check this experiment is five and make in to five different queries. After the query is provided to the system the similarity of the query with respect to the case are generated as shown in table 6.4 bellow.

Query	Description of Query	With respect to case	Degree of similarity
Query 1	The same value for all attributes	Case 3	1.0
Query 2	A value of attribute “nationality” is changed	Case 3	0.91
Query 3	A value of attribute “nationality” and “wereda” is changed	Case 3	0.84
Query 4	A value of attribute “nationality”, “Age” and “wereda” is changed	Case 3	0.79
Query 4	The same value for all attributes	Case 49	1.0
Query 5	A value of attribute “age” is changed	Case 49	0.89
Query 6	A value of attribute “age”, “Gender” and “nationality” is changed	Case 49	0.72
Query 7	The same value for all attributes	Case 175	1.0
Query 8	A value of attribute “travel purpose” is changed	Case 175	0.92
Query 9	A value of attribute “ travel purpose” & “ length of stay is changed	Case 175	0.71
Query 10	A value of attribute “travel purpose” “length of stay” “nationality” and “Gender” is changed	Case 175	0.64

Table 6.4: query similarity with their corresponding cases from the case base.

The case similarity test result of this experiment shows that when the test case has attributes value the same as a case stored in case base, the degree of similarity(global similarity) becomes 1.0(i.e. exact match) as in query 1, query 4, and query 7 as shown in table 6.4. On the other hand, the degree of similarity decreases when there is a change in one or more attribute values of the test case as compared to a case from the case base. When attribute values with higher weight value is changes the degree of similarity highly decreases.

6.4. Testing the CBR Cycles and Evaluating the Performance of the System

Now, this is the time to test the functionality of CBR cycles and the soundness of the prototype using selected test cases to check its validity and performance. The effectiveness of the prototype is measured with recall and precision using test cases. The evaluation of retrieval and reuse process of the system is presented as follows:

6.4.1. Evaluation of the Retrieval Process

For the purpose of this study, the effectiveness of the retrieval process of the recommender system is measured by using recall and precision. As stated by, McSherry (2001), recall and precision values are the most commonly used measures of the performance of the retrieval process in any CBR system. Recall measures the proportion of relevant cases to a given new cases (query) that have been retrieved from all the relevant cases in the case base. Precision on the other hand, measures the proportion of relevant cases to a given new case (query) from those that have been retrieved. Both recall and precision, being ratios, give values between 1 and 0.

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

$$\text{Recall} = \frac{\text{number of relevant cases retrieved}}{\text{Number of relevant cases in the case base}}$$

To begin this evaluation, for each test case the relevant visitor 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. Then the domain expert uses the value of recommendation attribute of visitor's cases as the main concept to assign the relevant cases to the queries, i.e. visitor cases that have similar solution (recommendation) are relevant to each other. Based on this concept, recall and precision are calculated.

Table 6.5 below shows sample test case with their corresponding relevant tourist cases that are assigned by the domain expert from the case base.

Test Case	**Relevant cases from the case base**
Case 364	Case 521, case 19, case 273, case 95, case 476, case 559, case 314, case 603, case 66, case 44
Case 277	Case 29, case 550, case 423, case 615, case 92, case 478, case 73, case
Case 472	Case 381, case 473, case 88, case 576, case 400, case 562, case 226, case 12
Case 44	Case 92, case 559, case 400, case 73, case 562, case 51, case 231, case 606, case 17
Case 556	Case 562, case 73, case 559, case 400, case 92, case 44, case 605, case 500
Case 600	Case 17, case, 43, case 604, case 605, case 606, case 20, case 99,
Case 12	Case 226, case 78, case 231, case 233, case 51, case 499, case 46, case 519
Case 311	Case 51, case 314, case 312, case 78, case 66, case 497, case 19, , case 364, case 13, case 607, case 2

Table 6.5: Relevant Cases Assigned by the Domain Expert for Sample Test Cases

Once the relevant cases are identified and assigned to the test cases the next step is calculating the recall and precision value of the retrieval performance of the CBR system with a threshold interval. As indicated in the research of Getachew (2012) and yibeltal 2013), there is no standard threshold for the degree of similarity that has been used for retrieving relevant cases in CBR. Different CBR researchers use different case similarity threshold. The above researchers used a threshold level of [1.0, 0.8) i.e. this means cases with global similarity score greater than 80% are retrieved. In this research, the threshold value is also set by the researcher. So, since the above researchers are satisfied in the selection of threshold value of [1.0, 0.8], for this research the researcher also used the threshold value of [1.0, 0.8]. The researcher conducted forty one (41) experiments to measure recall and precision by using a leave-one-out cross validation testing proportion and [1.0, 0.8] threshold interval.

After computing the similarity, the system generates several similar cases. Based on the similarity of cases here, the researcher calculates recall and precision value of the system assuming that the first 7 cases which are retrieved using 41 cases as a case base for experimentation are the best or most relevant one assuming that the first 14 cases are selected as a retrieved case in the experiment as shown in the following table (table 6.6).

Test case	Recall	Precision
Case 364	0.7	0.71
Case277	1.0	0.5
Case 472	0.87	0.57
Case 44	0.77	0.64
Case 556	0.87	0.57
Case 600	1.0	0.5
Case 12	0.87	0.57
Case 311	0.63	0.78
Average	0.83	0.61

Table 6.6 Performance Measurement of the system using Precision and Recall

As shown in table 6.6 above, recall of each test case can be calculated by dividing the number of relevant retrieved cases with total relevant cases. For example, for case 364, seven relevant cases are retrieved out of the total ten relevant cases in the case base. When we calculate its recall value, the result will be 7/10 which becomes 0.7 (70%). All the other test cases can be calculated in a similar fashion. The precision value of the system can also be calculated as the number of relevant cases retrieved divided by total retrieved cases. For instance, case 364 contains 7 relevant cases and total of 14 retrieved cases. Therefore the precision value of case 364 is 10/14 which is 0.71 (71%). The precision value of all the other test cases can be calculated in such a way. The average recall and precision value of the system is 83% and 61% respectively which is also a promising result. As shown in table 6.6 above, both recall and precision results are above average which is a good result. As seen in the table 6.6, for every test case more than average is registered both recall and precision. In terms of recall this research achieved a very good result. But, precision is somewhat lower compared to the average recall. This is because of the tradeoff between precision and recall and small number of cases. Recall in information retrieval is the fraction of the documents (case) that are relevant to the query that are successfully retrieved. It is the ability of a retrieval system to obtain all or most of the relevant documents in the collection (McSherry, 2001). The higher recall value shows that the system obtains most of relevant cases from the case base. Hence, this recommender system can retrieve relevant cases that enable visitors/tourists to make decision easily in attraction area selection process. On the contrary side, the system retrieved relevant cases to the user with 61% precision.

6.4.2. Evaluation of the Reuse process

The primary target of reuse process in this study is, to evaluate whether the proposed system recommend possible attraction areas and convenient time correctly for new tourist cases, or not 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 and it is defined by the percentage of the number of correctly recommended cases (McSherry, 2001).

To evaluate performance of the reuse process of case based recommender system in tourist attraction area, the researcher has been conducted 41 experiments and the result shows an accuracy which is above average and a hopeful result.

Total number of tested case query	Total number of correctly recommended cases	Accuracy in percentage
41	35	85.4%

Table 6.7: Accuracy value of the reuse process

6.4.3. Performance Comparison of the recommender system with Previous CBR Systems

The researcher tried to compare the performance of the system with previous research works on the area of CBR using recall and precision as a comparison criteria. But most of the CBR systems developed by previous researchers were focused on the retrieval task of cased based reasoning system by ignoring the reuse phase of the system.

The performance comparison of this case based recommender system using recall and precision with that of the previous works is depicted in the table 6.8 bellow:

Researcher Domain area	Programming tools Used	Retrieval task		Reuse task
		Recall	Precision	Accuracy
Getachew(2012), Mental Health	JCOLIBRI	82%	71%	Not evaluated
Alemu(2010) AIDS	JCOLIBRI	72%	63%	Not evaluated
Yibeltal(2013) Investment	JCOLIBRI	85%	64%	87%
Henok(2011) Hypertension	Python	86.1%	60%	88.89%
Biazen(2013) Field of study selection	JCOLIBRI	85%	55%	Not evaluated
Tourism(Tourist attraction area selection)	JCOLIBRI	83%	61%	85.4%

Table 6.8: a comparison of the system with the previous CBR systems

As shown in the table 6.8 above, the recall value of the system is nearly the same with the recall value of Getachew (2012). While the value of precision shows some improvement from the precision value of Biazen (2013) and almost the same as with the precision value of Henok (2011). On the other hand the accuracy of the reuse performance of the system registered somewhat lower than the reuse value of Yibeltal(2013), and Henok (2011). But the other researchers didn't evaluate the reuse value of their system.

Finally, when the researcher compares the results responded by domain experts and visitors, the performance of the case based recommender system responded by visitors 82% and domain experts 80% which is above average in acceptance of the system in users. This shows that the developed system is more acceptable and applicable in the domain area. Additionally, the testing procedure by using test cases helped to analyze the performance of the prototype

knowledge based system. The result obtained using test cases indicate that the prototype has recall performance of 83% and precision performance of 61%.

And also the accuracy of the prototype system for reuse process achieves 85.4%, which is above average. As the objective of reuse process in this research is to recommend correctly for visitor cases, i.e. to solve the problem correctly, the performance of the reuse process is measured by using accuracy. As a result of accuracy result the developed prototype system has the capability to advise and recommend in attraction area and appropriate visiting time selection correctly. The case similarity test result of this experiment shows that when the test case has attributes value the same as a case stored in case base, the degree of similarity(global similarity) becomes 1.0(i.e. exact match). On the other hand, the degree of similarity decreases when there is a change on one or more attributes value of the test case as compared to a case from the case base.

At the end, all the evaluation and testing results of the prototype shows encouraging finding for further research work to fully implement and apply case based recommender systems in recommending tourist attraction area and convenient visiting time in Ethiopia.

CHAPTER SEVEN

7. CONCLUSION AND RECOMMENDATIONS

7.1. Conclusion

Recommender systems have proven to be an important response to the problem of information overload by providing users with more proactive and personalized information services. CBR enables people to make 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 circumstances. CBR can work with new cases that match partially to the case from the case base as compared to rule based reasoning. Though, 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 a closed assumption where every fact are known and represented.

As studies shown that, the advising services given on the area of tourism in Ethiopia is in its infant stage. There are various factors that affect the tourism sector to be in its infant stage. Among these, shortage of skilled man power in the area, lack of guide line or criteria to assign visitors in different attraction area, absence of experts that can provides consistency advising service, and lack of awareness on the side of visitors about the purpose of advising systems for the selection of tourist attraction area and visiting time.

To tackle the above stated problems, the researcher initiated to conduct a research having the main goal of developing a prototype CBR system for the tourism sector that can assist the domain experts in assigning visitors in various attraction areas. The system aims to assist both the domain experts and visitors in the processes of making proper attraction areas selection decisions from already solved visitor cases.

Pertinent knowledge required for the development of the recommender system was acquainted from domain experts on the area, visitors and relevant documents through various mechanisms. In acquiring the knowledge the researcher used semi-structured

interview as a methodology. During prototype development, previous tourist cases were collected from NTO and MoCT. The pertinent knowledge acquired from various sources is conceptually modeled using hierarchical structure conceptual modeling technique. 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 of system is developed by using jCOLIBRI 1.1 Programming tool.

The CBR system uses the well-known CBR cycles (Retrieval, Reuse, Revise and Retain) to perform different tasks. In this recommender system, the first task is retrieval of cases by entering a new problem description (case) by using the query window. Next case similarity computation is performed and retrieves most similar cases. After retrieval of similar cases, reusing the previously solved cases from the case base is performed and followed by manual revision of cases to fit the problem at hand by tourism experts. The last task is storing the revised case in the case base for future use. Case base recommender system use past experiences as the domain knowledge and can often provide a reasonable solution through appropriate adaption. Since the system stores the new case within the existing cases, new case can be used as a case base for the next time. The retrieval task of the prototype used in this study is Nearest Neighbor retrieval algorithm.

Regarding to the evaluation process of the system, the recommender system registers encouraging retrieval performance which is an average value of 83% recall and 61% precision. Whereas its reuse performance scores an average value of 85.4% which is also a promising result. The system was also evaluated from the users' side which is called user acceptance testing. Then domain experts and visitors were involved for the evaluation and the average user acceptance evaluation registered 80% and 82% performance by domain experts and visitors respectively.

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

- The major attribute that have more influence in tourist attraction area and visiting time selection are: age, gender, location of attraction areas (such as region , zone, woreda), nationality, length of stay ,travel purpose and visitors interested attraction area.

- The applicability of case based recommender system for tourist attraction area and visiting time selection haven been proved.

- 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 attraction area and visiting time to visitors.

- In the proposed case based recommender system learning is made for new cases by dynamically updating in the existing case base for the purpose of future use as a case base.

- The proposed system contributes a lot in the selection or recommendation of attraction area and visiting time by assisting tourism experts and visitors in the domain area. The proposed prototype system actually contributes a lot, especially for those less experienced experts or in the absence of expert.

7.2. Recommendations

Even though, promising results are observed under this study, there are problem areas that need further investigation for future work. Therefore, the researcher recommends the following issues as a future research direction based on the findings of this study.

- The retrieval algorithm used for retrieval of cases for the application of this system is nearest neighbor retrieval algorithm. Since the case base of the system increases through incremental learning, the retrieval time increases linearly. Therefore, the retrieval performance will decrease from time to time. To overcome this problem investigating case maintenance techniques is essential and in the future it is recommended to use other retrieval algorithms such as template retrieval that returns all cases that fits certain parameter.
- The relevance of the system in the domain area is already proved by measuring its performance. But the system was developed in English language and is difficult to understand by some visitors. Further investigation can be conducted by developing a case based recommender system in different local languages. This helps visitors to communicate using their own language with the case based recommender system.
- This recommender system recommends only one recommendation even if other cases which are similar to the new case are displayed in ranked order. For the future work, the researcher recommended to use other tools that can recommend more alternative solutions.
- The relevant attributes used for this research are collected from the previous tourist cases from NTO and MoCT. These attributes are not sufficient for the selection of attraction area and visiting time decision. So further research can be conducted by adding other important attributes such as housing preference, level of education, marital status, and purchasing habits by making a direct survey of successful visitors.
- The explanation facility of the developed system does not give response or feedback based on visitor's questions. It gives only a direct explanation when the system recommends a solution. The explanation facility gives explanation only about the

recommended (assigned) attraction area, but does not give it at any time the user needs explanation. So further research can be done to add explanation facility that can give clarification any time the user wants it in addition to explanation of the recommended attraction area.

- The performance of the system can be improved if hybrid approach is employed by combining rules, cases and models since these rules, cases and models have complementary strength. The inclusion of rule based reasoning in this research helps the proposed system to give explanation facility when the user wants explanation and used to represented fact and rules extracted from the domain expert. Therefore, for the future, it is better to integrate these approaches to make knowledge base system more successful.

Reference

1. Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, 7(1), pp. 39-59.
2. Akerkar, R., & Sajja, P. (2010). Knowledge-based systems. Jones & Bartlett Publishers. 1–11.
3. Bergmann, R. (1998). Introduction to case-based reasoning. Department of Computer Science, University of Kaiserslautern.
4. Bergmann, R., Kolodner, J., & Plaza, E. (2005). Representation in case-based reasoning. The Knowledge Engineering Review, 20(3), 209-213.
5. Breuker, J. & Wielinga, B. (1985). KADS: Structured Knowledge Acquisition for Expert Systems. Proceedings of Expert Systems and their Applications, Avignon.
6. Briggs, P. (2009). Ethiopia. Bradt Travel Guides (5th ed.). Bradt Travel Guides Ltd. Bucks, England.
7. Burge, J. E. (2001). Knowledge elicitation tool classification. Artificial Intelligence Research Group, Worcester Polytechnic Institute.
8. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4), 331-370.
9. Burke, R. (2006). Knowledge based recommender systems, University of California, Irvine.
10. Burke, R. (2007). Hybrid web recommender systems. In The adaptive web, 377-408. Springer Berlin Heidelberg.
11. Cooke, N. (1994). Varieties of knowledge elicitation techniques. International Journal of Human-Computer Studies, 41(6), 801-849.
12. Corniel, M., Gil, F., Molero, J., Ferrer, J., Borges, A., Gil, R., & Contreras, L. (2011). Studies orientation and recommendation system (SORS): use case model and requirements. NAUN International Journal of Systems Applications, Engineering & Development, 3, 387-395.
13. Culture and tourism office (2011). Tourism Development Strategy, Addis Ababa, Ethiopia.
14. Cunningham, P., & Zenobi, G. (2001). Case representation issues for case-based reasoning from ensemble research. In Case-Based Reasoning Research and Development, Springer Berlin Heidelberg, 146-157.

15. Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation?. AI magazine, 14(1), 17.
16. De Gemmis, M., Iaquinta, L., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2009). Preference learning in recommender systems. Preference Learning, 41.
17. Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. International journal of information management, 20(1), 17-28.
18. Ejigu, T. (2012). Developing knowledge based system for cereal crop diagnosis and treatment: the case of kulumsa agriculture research center, Addis Ababa University, Ethiopia.
19. Ethio – American Trade and Investment Counsel, Facilitating trade and Investment between Ethiopia and The USA. Thirteen Months of Sunshine. Address: http://www.eatic.org/sunshine.html Accessed 1, January, 2013.
20. Ethiopia T., (2002). Application of Case-based Reasoning for Amharic Legal Precedent Retrieval: A Case Study with the Ethiopian Labor Law. MSc Thesis, School of Information Science, Addis Ababa University, Ethiopia.
21. Fabiana, P. Lorenzi, F., & Ricci, F. (2005). Case-based recommender systems: a unifying view. In Intelligent Techniques for Web Personalization (pp. 89-113). Springer Berlin Heidelberg.
22. Getachew W. (2012). Application of case-based reasoning for anxiety disorder diagnosis. MSc Thesis, School of Information Science, Addis Ababa University, Ethiopia.
23. Ghauth, K. I. B., & Abdullah, N. A. (2011). The Effect of Incorporating Good Learners' Ratings in e-Learning Content-based Recommender System. Educational Technology & Society, 14(2), 248-257.
24. Guy, I., Zwerdling, N., Ronen, I., Carmel, D., & Uziel, E. (2010).Evaluating Recommendation Systems: Introduction to Recommender System.
25. Haffner, E., Heuer, A., Roth, U., Engel, T., & Meinel, C. (2000). Link Proposals with Case-Based Reasoning Techniques. In WebNet World Conference on the WWW and Internet, 2000(1), 233-239.
26. Heines, J. (1983). Basic concepts in knowledge-based systems. Machine Mediated Learning, 1(1), 65-95.

27. Janet E. Burge, (1997). Knowledge Elicitation Tool Classification, Artificial Intelligence Research Group Worcester Polytechnic Institute.
28. Jönsson, C., & Devonish, D. (2008). Does nationality, gender, and age affect travel motivation? A case of visitors to the Caribbean island of Barbados. Journal of Travel & Tourism Marketing, 25(3-4), 398-408.
29. Juan A., Recio, J., Sánchez, A., Díaz-Agudo, B., & González-Calero, P. (2005). jCOLIBRI 1.0 in a nutshell. A software tool for designing CBR systems. In Proceedings of the 10th UK Workshop on Case Based Reasoning, 20-29.
30. Kolodner, J. (1992). An introduction to case-based reasoning. Artificial Intelligence Review, 6(1), 3-34.
31. Kriegsman, M., & Barletta, R. (1993). Building a case-based help desk application. IEEE Intelligent Systems, 8(6), 18-26.
32. Lambadina Tour & Travel (2009). Message from Lambadina. Address: http://www.lambatour.com/php/home.php. Accessed: 28 December, 2013.
33. Main, J., Dillon, T. S., & Shiu, S. C. (2001). A tutorial on case based reasoning. In Soft computing in case based reasoning (pp. 1-28). Springer London.
34. Massa, P., & Avesani, P. (2007, October). Trust-aware recommender systems. In Proceedings of the 2007 ACM conference on Recommender systems (pp. 17-24). ACM.
35. McSherry, D. (2001). Precision and recall in interactive case-based reasoning. In Case-Based Reasoning Research and Development Springer. Berlin Heidelberg, 392-406.
36. Ministry of culture and tourism (2012). Manuals of Ethiopian tourism guide, Addis Ababa, Ethiopia.
37. Ministry of urban development and construction (2009). Urban tourism and heritage planning manual, Addis Ababa, Ethiopia.
38. Motta, D., Christopher A., & Sheldon H. (2009). Knowledge Acquisition as a Process of Model Refinement, Human Cognition Research Laboratory. The Open University.
39. Muhammad, H. (2006). Evaluation of jCOLIBRI, a Master's thesis report, Malardalen University, Vasagatan 44 72123 Vasteras, Sweden.
40. Pamela, H. (2009). Tourism and Hospitality studies theme parks and attraction.

41. Pasquale, M., Lops, P., De Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In Recommender systems handbook (pp. 73-105). Springer US.
42. Perugini, S., Gonçalves, M., & Fox, E. (2004). Recommender systems research: A connection-centric survey. Journal of Intelligent Information Systems, 23(2), 107-143.
43. Prentzas, J., & Hatzilygeroudis, I. (2007). Categorizing approaches combining rule-based and case-based reasoning. Expert Systems, 24(2), 97-122.
44. Ranjan, K., Frank, W. & Nasuti, P. (2006). Knowledge Acquisition, Representation, and Reasoning. 69–79.
45. Recio-García, A., Díaz-Agudo, B., & González-Calero, P. (2008). jCOLIBRI2 Tutorial Document version 1.2. Group for Artificial Intelligence Applications Universidad Complutense De Madrid.
46. Robin Burke, (2007). Hybrid Web Recommender Systems, School of Computer Science, Telecommunications and Information Systems DePaul University.
47. Sagheb-Tehrani, M. (2009). A Conceptual Model of Knowledge Elicitation. In Proceedings of Conference on Information Systems Applied Research (CONISAR) 2009 (Vol. 1542, pp. 1-7).
48. Salem, P., Tollrurst, D., & Emmerson, S. (2005). A Case Base Experts System for Diagnosis of Heart Disease. International Journal on Artificial Intelligence and Machine Learning, 5(1), pp. 33-39.
49. Santos, O., & Boticario, J. (2011). Requirements for semantic educational recommender systems in formal e-learning scenarios. Algorithms, 4(2), 131-154.
50. Schafer, J., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In The adaptive web (pp. 291-324). Springer Berlin Heidelberg.
51. Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W., & Wielinga, B. (1999). Knowledge Engineering and Management: the CommonKADS Methodology. A Bradford Book.
52. Shimazu, H. (2002). ExpertClerk: A Conversational Case-Based Reasoning Tool for Developing Salesclerk Agents in E-Commerce Webshops. Artificial Intelligence Review, 18(3-4), 223-244.

53. Tagel, A. (2013). Knowledge based system for pre medical triage treatment at Adama university Asella hospital. MSc Thesis, School of Information Science, Addis Ababa University, Ethiopia.
54. Toffler, A. (1981). The third wave (pp. 32-33). New York: Bantam books Ltd.
55. Tourism Development Policy (2009). Federal Democratic Republic of Ethiopia Tourism Policy, Ministry of Culture and Tourism August 2009 Addis Ababa, Ethiopia.
56. Tourism Ethiopia the Historic Route (nd). Sample Itineraries of Suggested Routs. Address:http://www.tourismethiopia.org/pages/historic-route.asp#BGLA. Accessed 27 December, 2013.
57. Tutu J. (2011). International tourism marketing - promoting BRC budget car rental and tour, Ethiopia.
58. United Nations World Tourism Commission (2007). Tourism Highlights. 2007 Edition.
59. Vignette, M., (2004). Knowledge Acquisition, Representation, and Reasoning, United Kingdom.
60. Von Wangenheim, C. (2000). Case-Based Reasoning–A Short Introduction. Universidade do Vale do Itajai.
61. Wang, W., Ting, S., Tse, Y., & Ip, W. (2011). Knowledge elicitation approach in enhancing tacit knowledge sharing. Industrial Management & Data Systems, 111(7), 1039-1064.
62. World Bank (2006). "Ethiopia: Towards a Strategy for Pro-Poor Tourism Development", the World Bank Private Sector Development Country Department for Ethiopia.
63. www.tourismethiopia.org. (Official website of Ministry of Culture and Tourism of the Federal Democratic Republic of Ethiopia)
64. Yabibal, M. (2010). Tourist Flows and Its Determinants in Ethiopia. Ethiopian Development Research Institute Addis Ababa, Ethiopia.
65. Yechale, M. (2011). Tourism certification as a tool for promoting sustainability in the Ethiopian tourism industry. Addis Ababa University, Addis Ababa, Ethiopia.
66. Yibeltal, C. (2013). Application of recommender system in investment sector. MSc Thesis, School of Information Science, Addis Ababa University, Ethiopia.

Appendixes

Appendix I

Interview questions to Domain Experts

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. What type of advising system is given when Tourists come to visit?
2. What are the criteria to consider assigning Tourists in different Tourist attraction areas?
3. What attributes are given more consideration in order to identify Tourist attraction areas that best mach with the Tourists?
4. Which Tourist attraction areas are more preferable to female Tourists and which one is for male Tourists?
5. What is the influence of Tourist attraction area location to be success /to be satisfied in recreational program?
6. Which Tourist attraction areas are reserved for government, reserved for domestic Visitor, reserved for foreign visitor?
7. What are the main reason for majority of visitors are not successful/ satisfied?

Appendix II

Interview questions to Visitors

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. How to get advising systems from domain experts

2. What are the problems in advising systems of NTO &MOCT?

3. Is there any experienced Tourism experts that gives a brief advice on how to visit and where to visit?

4. How to select the tourist attraction areas?

5. Have you get any guidance from experts to select tourist attraction areas?

Appendix III

Prototype Evaluation form for the Domain Expert and visitors This is an evaluation form to be filled by tourism experts in order to evaluate the applicability of the case-based recommender system in tourist attraction area and visiting time 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 tourist attraction area selection.

No	Evaluation criteria	Performance value						
		1	2	3	4	5	Average	%
1	Easy to use of the recommender system							
2	Is the system efficient in time							
3	Is the user interface interactive							
4	Adequacy and clarity of decision support							
5	Relevancy of the retrieved case in the decision making							
6	Fitness of the final solution to the new case							
7	Relevancy of the attributes in representing visitors case							
8	Does the explanation facility give brief description about the recommended attraction area							
9	Rate the significance of the system in the domain area							

DECLARATION

This thesis is my original work. It has not been submitted for a degree in any other university and all sources of material used for the thesis have been duly acknowledged.

_____

Tamir Anteneh Alemu

June, 2014

_____

Advisor: Rahel Bekele (phD)

June, 2014