Concept based term weighting (CBW) in online second hand car trading

Said Mohammad Al Tahat

University Utara Malaysia

2009

Concept based term weighting (CBW) in online second hand car trading

A thesis submitted to the College of Arts and Sciences in Partial in Fulfillment of the requirement for the degree Master of Science (Information Technology)

Universiti Utara Malaysia

October 2009

By

Said Mohammad Al Tahat

KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

SAID MOHAMMAD ALTAHAT
(801618)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**

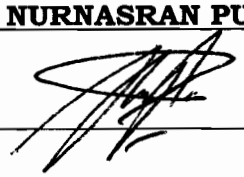
telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

CONCEP BASED TERM WEIGHTING (CBW) IN ONLINE SECOND HAND CAR TRADING

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **MR. NURNASRAN PUTEH**

Tandatangan
(Signature) : 

Tarikh
(Date) : 16 Nov 2009

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor(s) or, in their absence, by the Dean of the Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Request for permission to copy or to make other use of materials in this thesis, in Whole or in part, should be addressed to:

Dean of the Graduate School
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman

ABSTRACT

With the increasing of life demands and challenge people face problems in buy or sale second hand car. Therefore online second hand car by using internet technology is proposed to minimize such challenges and demands, by allowing people to buy or sale car more efficiently and easily. The objective of this work is two-fold; to combine information on online second hand car services from many second hand car into a single system and to use a weighting scheme to prioritize a customer's preferences on the car to be buy or sale The proposed approach is intended help customers in sale or buy the required car from second hand car trading that offers good price.

ACKNOWLEDGME NTS

"In the Name of Allah the Most Gracious and Most Merciful"

First and foremost, I am forever indebted and thankful to Allah for his blessings without which the whole would not have been possible.

I will always be especially thankful to my supervisor Mr. Nurnasran, for his guidance, patience and faith in me, he was always positive, supportive and encouraging, he fostered my academic growth by challenging and inspiring me to each deeper, to learn more, to expand my viewpoint, and to think critically. Yet, he allowed me to express my views openly and to disagree even when I was wrong. Also I would like to thanks all members of CAS college.

The final and the most important acknowledgment are owed to my family specially to my brothers Ahmad, Dr. Yousef and Ali to the source of my light and pleasure to the one who enlightens my life, to my dear mother, to all my sisters and all my brothers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

This chapter represents the background of the study, problem statement, and research questions, objectives of the study and scope and limitation of the study. It provides a description to the undertaken study.

1.2 BACKGROUND TO THE STUDY

Information retrieval (IR) is the science of searching for text, for information within text and for metadata about text, as well as that of searching the World Wide Web and relational databases. It also partly covers the usage of the terms data retrieval, document retrieval, information retrieval, and text retrieval, but each also has its own corpse of literature, theory, praxis and technologies. The detailed information about this topic will be discussed further in literature review in chapter two.

The importance of hypertext is growing during the last decade. Information systems and the Internet uses hypertext format where the data is organized associatively rather than sequentially or relationally. In contrast to regular text, hypertext has a non-linear structure and the techniques of pattern matching for text cannot be directly applied to hypertext (Amir et al., 1997).

The critical idea carried out by a weighted based algorithm is that, when one attributes links to another, it is basically casting a vote for the other attribute. The higher number of votes that are cast for an attribute the higher significance of the attribute.

Tourism in Malaysia has been found to expand rapidly in the recent years (Mansor, 2007). In order for the tourist and also the local people to better explore Malaysia and perform any required journey, the transport facility is required. This includes the use of public transport and private car. And not all people can have new car. More people always try to buy second hand car. As buy or sell a second hand car may be a better option to some people, services offered by the second hand car trading opens various business opportunities.

In the conventional method, a potential customer of a second hand car trading would personally go the premise and identify the desired vehicle to be buying. Such an approach requires extra time, effort and money. To facilitate customer, most of the existing company have offered their services through websites, for example the (www.mudah.my) and (www.carlist.my). These websites offers a diversified range of services that include buy and sell car.

The related information on the second hand car trading (for example car model and car option) is presented in the company's website. Customers can visit this website and make queries about services offered by the company, such as the company's branches and the kinds of cars and models that available to sell. Nevertheless, these online second hand car trading do not provide the opportunity for customer to prioritize their personal preferences in buy the describe vehicle, personal preferences that the company can provide to the customer, for example in the (www.mudah.my) website a customer could not indicate which of his requirement is the most important factor to be considered in buy a car. And can't allow customer to sell has care.

Thus, the focus of this study is to apply the weighed-based approach in identifying and retrieving the required information. That is, find a good and reliable car, better services and best price from different second hand car trading. Even though customers can do that by visiting the relevant second hand car trading premises personally and present their requirements such an activity will take a lot of time, money and effort that the customer may then end up hiring a car from the first company he/she has visited.

1.3 PROBLEM STATEMENT

The existing system in the second hand car, such as (www.mudah.my) and (www.carlist.my) do not provide customers with the opportunity to include their requirement in buying or get high price in sale. Such a facility would allow a customer to prioritize his/her preferences such as type of car, price, car model, and car option, in identifying the suitable car to buy or get good price to sell the car. Furthermore, in order for a customer to get the best price, he/she has to personally visit second hand car websites or read the adverts on the newspaper. Such an approach is time consuming and a customer might be put it off and proceed with the information in-hand.

In order to overcome these problems, we adopted the concept based term weighting (CBW) to prioritize user preferences in identifying and retrieving the required information from several resources. If the customer wants to sell a car, he/she will enter the full information about his/her car and the system will estimate the car price. But, if the customers need to buy, he/she will enter his/her preferences and the system will search in several databases and best matches with the users references.

1.4 RESEARCH QUESTIONS

The questions that need to be answered in this work are:

1. How the concept based term weighting (CBW) can be adapted in a second hand car application?
2. What are the characteristics in CBW to be included in second hand car trading application?

1.5 OBJECTIVES OF THE STUDY

This study has two objectives:

1. To develop a second hand car system that uses a Concept based term weighting (CBW) in matching user preferences.
2. To evaluate the proposed approach (objective 1) by comparing the acceptance of the developed application using Concept based term weighting with one using existing approach (hyper-text matching analysis).

1.6 SCOPE AND LIMITATION OF THE STUDY

The scope of this project is limited to retrieving information on second hand car from many different second hand cars in Malaysia. User preferences are limited to the following:

1. Kind of car.
2. Car model
3. Car color.
4. Year of the car
5. Car system e.g. Airbag.
6. Air-condition.
7. Engine capacity.
8. Type of transmission (Manual / Auto)
9. Kind of car body (Saloon, station, coupe)

1.7 ORGANIZATION OF THE REPORT

This report is organized as follows. Chapter 2 provides a general description on Web-Based second hand car trading for two companies background, in later chapters and discusses why the researcher believes that the proposed framework is suitable for Web-Based for traditional application problems .Chapter 3 discusses the methodology used in this study. Chapters four focus on evaluating the system and also present the result analysis. Finally Chapter 5 discussion the conclusion, results and the future work.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Today is the world of Internet. World is a global village. Internet has brought a lot of differences in the living style and specially shopping style. There is an increase in statics in the many of WWW for shopping purpose. The materialization of world-wide-web has shown the way to an augment concentration in means for searching information. An online second hand car trading web site is an essential part of life these days as to facilitate customer to buy or sale their required vehicle. Customers need an online method to find good and cheap cars to buy. Therefore, companies has uploaded their database online to ease customer regarding buying the cars. But there are problems in the existing online second hand car trading web site. The main problem is that the search is based on a single database and this generates limited choices for the user. Furthermore, user preferences are not represented in the existing application. Mostly are using local databases or if they are using the centralized database, there is redundancy in the database.

In this chapter, I try to collect the right information by using the right algorithm which can be integration with the requirement of this system, to help the researcher to choose the suitable algorithms that can reduce the user time to get what he need from the website , and to provide him by the accuracy information's that he need, I discusses several related studies and various search algorithm used for developing online second hand car trading web site, for two second hand car web site named ww.mudah.my, and www.carlasit.com. Various types of search are used to locate information either from the internet and database or directories. In this chapter we will begin with related works on Information Retrieval, search engine, weighted-based application and concept-based term weighting (CBW).

2.2 INFORMATION RETRIEVAL

According to Dietinger et.al (1999), make searching in the WWW and access to the information not mean you have retrieved the information of your use and that satisfy users' needs. The face problem is obtaining the right information which is needed with proper quality, reliability and timeliness and to get only information that has been requested: we will call this 'knowledge'. This can be achieved by clustering the data according to the properties of data/information. Examples of clustering applications are finding the uniform sub-populations or classification of sub-categories. Our work is based on this type of information retrieval where we intend to retrieve 'specific' documents based on a list of requirement.

Thomas, Dietinger et.al (1999), has proposed a solution named "Knowledge gathering process". The difference between the dynamic and static part is based on how we have the control and influence over the static part whereas our influence is least in the dynamic component of the background library. They stated that meta-data are very important for the classification and organization of knowledge as well as for finding relevant information by end users.

Krovetz & Croft (1992), reported on an analysis of lexical ambiguity in information retrieval test collections. They performed experiments to determine the utility of word meanings for separating relevant from non relevant documents. The experiments show that there is considerable ambiguity even in a specialized database. Word senses provide a significant separation between relevant and non relevant documents, but several factors contribute to determining whether disambiguation will make an improvement in performance. For example, resolving lexical ambiguity was found to have little impact on retrieval effectiveness for documents that have many words in common with the query.

As described in Hassan (2005), proposed heuristic mix is the most intelligent, user-friendly matching algorithms applied in IR and this algorithm tries to find a new operator by using the logical approach. He/she has proposed mixture of four methods for matching strings; from left to right, from right to left, in specific order and in any order.

- **Left to right matching:**

The search with deterministic finite automaton performs exactly N text characters inspections but it requires an extra space O(mJ). The forward Dawg Matching Algorightm performs exactly the same number of text characters inspections using suffix automaton of the pattern (Crochemore & Rytter, 1994).

- **Right to left matching:**

The Boyer-Moore algorithm is considered as the most efficient string-matching algorithm in usual applications. A simplified version of it (or the entire algorithm) is often implemented in text editors for the "search" and "substitute" commands. Cole proved that the maximum number of character comparisons is tightly bounded by 3n after the pre-processing for non-periodic patterns. It has a quadratic worst-case time for periodic patterns (Boyer & Moore, 2003).

- **In specific order matching:**

The algorithms of Colussi and Galil-Giancarlo partition the set of pattern positions into two subsets. They first search for the pattern characters, whose positions are in the first subset from left to right and then if no mismatch occurs they search for the remaining characters from left to right. The Colussi algorithm is an improvement over the Knuth-Morris-Pratt algorithm and performs at most 3/2 n text character comparisons in the worst case. The Galil-Giancarlo algorithm improves the Colussi algorithm in one special case, which enables it to perform at most 4/3 n text character comparisons in the worst case (Colussi & Fastest, 1994).

- **And in any order Matching:**

The Horspool algorithm is a variant of the Boyer-Moore algorithm, it uses one of his shift functions and the order in which the text character comparisons are performed is irrelevant.

This is also true for all the other variants such as the Quick Search of Sunday, Tuned Boyer- Moore of Hume and Sunday, the Smith algorithm and the Raita algorithm (Horspool, 1980)

According to Hassan (2005), Mix Heuristic Algorithm has relied on strict String Matching first and if the 100% solution (exact match) is not bring into being it uses near string matching to calculate the possibility of achievement and most possible solution. Exact string match is a bit challenging, principally when an accurate match is not found and user needs to change his explore by using various new explore strategies such as by minimising the search string another time and over again or some times with advance search query. If the user is new to the system and unaware of search strategies, he/she will not discover the relevant results for example user might not find what he/she exactly wants or he/she may find the results of no importance to him, now we should imagine that Mix Heuristic Algorithm is not only fast but also intelligent and user friendly.

According to Toms (2000) some time people find the information without looking for it through chance discovery, Information retrieval in traditional information systems is based on the idea that users know or incompletely know a little bit about the information they try to find. Currently, there are three

behaviours in which people get information, first one from the investigation for information about a well-defined and known thing(s), it mean the relationship is predetermined by the system's reply to the question, second one From the investigation for information about an object that cannot be completely make clear, but will be known on view, and 'third' from the chance discovery of an object. Here chance of information retrieval happens when a user with no a priori purposes interacts with a node of information and gets positive information. We can say that the retrieval was causes 'by chance' or a chance met (Toms, 1998b).

As described in Tomassen (2006), a difficulty through customary information retrieval (IR) systems is that they normally retrieve information with no clearly explained area of interest near the user. ontologies are used to enhance user-experience by getting the queries nearer to the user's needs. Improving seek result is excellence for extensive seek systems.

To describe how these systems get user's needs, refer to Figure 2.1.

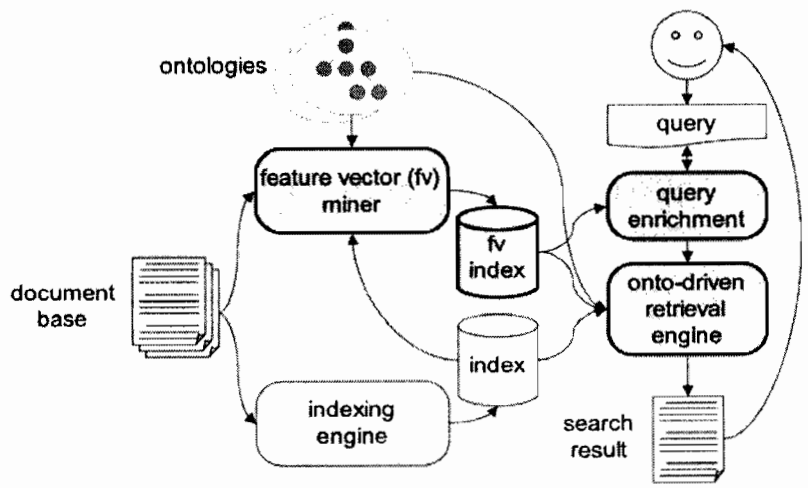


Figure 2.1: The Overall Architecture of the Ontology-driven Information Retrieval system (Source: Tomassen, 2006)

Figure 2.1, explains about ontology-driven information retrieval system. And to know the individual components, the fv index is created offline equal to the index of the search engine and feature vector used to get the related idea and take it out from gathering some documents. Another one is Indexing data to document collection and some time the query will be not clear then query enrichment will handle it. And the onto-driven retrieval engine used to carry out the search result and give it to the user as well as good result.

According to Lopresi (1996), the outcome from Optical character recognition the text is important from the perspective of retrieval, especially to address the issue of noise, the researcher combine two classic models for dealing with inexact data: approximate string matching and fuzzy logic. When building or improving models and algorithms used for information retrieval, it is frequently convenient to summarise that the all that is inside of the database are, in logic, perfect and understood: the text has been cautiously edited and proof-read, all of it suit, there is nothing not present;, and there are little if some errors.

An inconclusive study was carried out as described in Roberto & Velardi (2003) on query expansion which never proved effective except for the so-called "open domain question answering" task. They provided interesting evidence suggesting new guidelines for future research. Word sense disambiguation is in fact only one of the problems involved with sense based query expansion. The second is how to use sense information (and ontologies in general) to expand the query. They showed that expanding with synonyms or hyperonyms has a limited effect on web information retrieval performance, while other types of semantic

information derivable from ontology are much more effective at improving search results. Therefore they developed an algorithm that may be tuned to produce high precision, possibly at the price of low recall. The algorithm belongs to the class of structural pattern recognition methods. Structural pattern recognition is particularly useful when instances have an inherent, identifiable organization, which is not captured by feature vectors. In this work they use a graph representation to describe instances (word senses).

There is a technique named "Stemming". Stemming words to (usually) eradicate suffixes has applications in text search, machine translation, document summarization, and text classification. The concept is taken from the trees' stem. For example, English stemming reduces the words "introduction," "introducing" and "introduced" to their common morphological root, "introduc-." In text search, this permits a search for "introduction" to find documents containing all words with the stem "introduc-." (Xapian, 2007).

According to Zeinalipour et al. (2004), the peer to peer information-retrieval (IR) environment assumes that each peer has a database of documents (car, video, book, or other semistructured documents) that is shared in the network. A node searches for information by sending Query messages that contain sets of keywords to its peers. A peer receiving a query message compares the similarity of the query against its document collection. Typically, this involves finding documents that contain the set of query keywords. If the matching is successful, the peer generates a reply message that contains pointers to the matching documents. IR is more complex than traditional search techniques based on

object identifiers or filenames, because there is no central repository, there are large numbers of documents, and nodes join and leave in a dynamic and ad hoc manner.

Collier and Arnold (2003), presented a white paper about what might be called "classic search" for the degreed information professional, classic means Boolean queries passed against the collection of content unified by an editorial policy. Also he/she presented a schematic and simplified system architecture for a spider- based search and retrieval system dealing with the body of content (document).

In string searching, breadth search algorithm is better than A* and DFS because it uses character by character search to search a word and then string. This can be thought of as being like Dijkstra's algorithm for shortest paths (Black, 2006; Mosteo, 2006), but with every edge having the same length. However it is a lot simpler and doesn't need any data structures. We keep a tree (the breadth first search tree), a list of nodes to be added to the tree, and markings (Boolean variables) on the vertices to tell whether they are in the tree or list (Russell, 1995).

As described in Zhao & Karypis (2002), wording categorization and organization have been broadly studied in both information retrieval and wording mining literatures. Nevertheless, there is small work done on combining the category information with traditional Information Retrieval techniques to develop retrieval. The authors explore another way to make use of category information

by adjusting period weights based upon the term's distribution in the middle of categories. They presented normalized entropy (NE) method to settle on the category specificity of each term, from which they derived two supervised term weighting schemes. The evaluation results on TREC datasets give the idea that the proposed schemes outperform the traditional Intermediate Distribution Frame scheme significantly when the queries contain more than a only some specific terms and get competitive results on short and well-defined queries.

According to Riloff & Hollaar (1996), intelligent information retrieval is an awesome and amazing latest way for information retrieval research, which can generate more contented text image from a large and mixed information databases. Question IR systems have to manage in real time; we should be clever to search big numbers of text speedily and professionally. Other information-retrieval applications, such as text categorization, text routing, and text filtering, become more and more important. For example, a text categorization system for one word or one sentence might use categories Information-retrieval systems have to wrestle with all of the ambiguities in normal language, such as synonymy (e.g., "end", "finish", and "stop" have roughly a same meaning) and polysemy (e.g., "get" has many different meanings, for example some time we use it rather than win, bring, and request). Phrases also need good and special thought because multiword expressions sometime have a composite meaning different from the individual words.

2.3 SEARCH ENGINE

Google's technology makes an analysis for complete content of a page and factors in fonts, subdivisions as well as the precise location of each word. Also makes an analysis on the content of nearest web pages to make sure the outcome returned are the most related to a guests query. This is what we are trying to achieve as we intend to present user with the most relevant documents as depicted in their user preferences.

As describe in Brin & Page (2000), Google employs a number of techniques to improve search quality including page rank, proximity information and Anchor text. Page rank is an excellent way to prioritize the results of web keyword searches and helps a great deal; proximity information helps increase relevance a great deal for many queries, and the third one is Anchor text it help us to search non-text information, often provide more accurate descriptions of web pages than the pages themselves. Also anchors might exist for documents which cannot be indexed by a text-based search engine, for example the images, databases, and programs. This makes it probable to come back web pages which have not actually been crawled.

According to many studies such as Marchiori (1997), Spertus (1997), Weiss (1996) and Kleinberg (1998), there is quite a bit of recent optimism that the use of more hypertextual information can help improve search and other applications (as cited from Brin & Page, 2000)

According to Craswell (2000), a search broker acts as an intermediary between a user searching for information and a set of search servers. It may perform automatic server selection, choosing servers which are likely to be most useful. It may also concurrently query the selected servers and present their results to the user in a single merged list. The effectiveness of a broker over a given set of servers depends on the effectiveness of its server selection and results merging methods. Its selection method must choose servers which return relevant documents. Its merging method must rank the combined results. This is similar to what we are trying to achieve in this study. We intend to combine technical result from three databases and later sort the results in order to fulfil user needs.

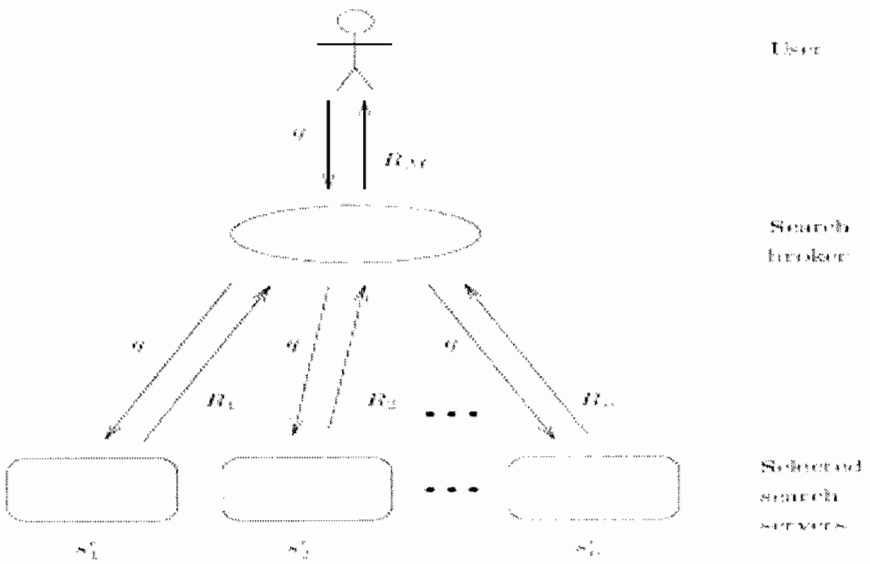


Figure 2.2: Search broker Network communication (Craswell 2000)

In a search broker network, (Figure 2.2), the clients first supply the broker with a query q, also throughout the broker's graphical client interface or by linking to the broker using a client such as a Web browser. The broker chooses servers ($s1$,

*s*2. . . *sn*) retrieves their outcomes (*R*1, *R2*.... *Rn*) and produced a combined results list *RM* to clients.

And another search broker is a search broker information flow, it Given *S* and *q* the broker selects a subset *S*0 _ *S*, retrieves *R*1, *R2, R3* from those servers and makes the combined list *RM* to the clients as follows:

$$\langle S, q \rangle \xrightarrow{Selection} \langle S', q \rangle \xrightarrow{Retrieval} \langle (R_1, \ldots, R_{|S'|}), q \rangle \xrightarrow{Merging} R_M$$

As described in Yuwono & Lee (1996), Figure 2.3, show the universal architecture of search engine, where an index server covers a number of Would Wide Web HTML document on the Internet belonging to a cluster based on some categories as a institution.

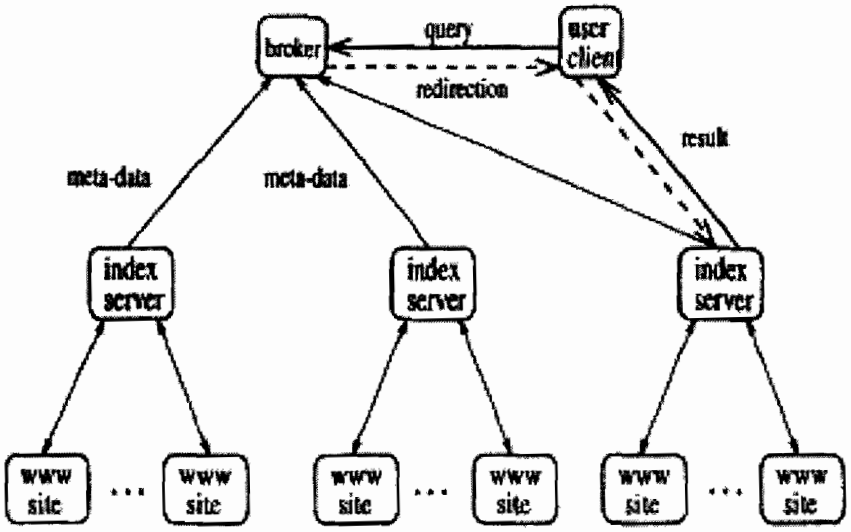


Figure 2.3: Multiserver Organization and Query Routing

(Source: Yuwono & Lee, 1996)

World Wide Web server received a question from a client and gave them back as a list of Uniform Resource Locators ranked by state of being connected to the current score.

According to Deogun et al. (1998), World Wide Web have a lot of information and mostly we feel that it suggest and offers confused environment Web guests to get information because the explanation of text by HTML is understandable by human, but is not so by machines. Researchers assume that information is suitably structured to make easy helpful Web mining. Also they don't give advertence to this matter, but rather investigate the performance of a search engine while indexing more logical elements of HTML documents and while increasing the scope of indexing process.

(Deogun, et al. 1998), measure the quality of retrieval output by MZ metric. Assume at least have one written file is related to a consumer question. It is clarify by the formula:

MZ (a, b, c) = 1- a/ a + b + c,

where a is number of relevant & retrieved text, b is number of nonrelevant & retrieved text, c is a number of relevant & not retrieved text. We should know that the good retrieval corresponds to lower MZ value, which varies among zero and one.

Most important steps of the experiment are select a Web list, decide a set of questions, let a set of structural abstractions, denoted by Sstructures; be {{Table}; {Table; list}; {Table; list; paragraph}} and I index variable on such

set. let I = 0, Let the set of the search region, denoted by Sregion; be {r; r with s; r with s with m}; where r, s, and m are set of referential, special, and miscellaneous pages.Let j be index variable on Sregion: Let j = 0, filter out the structures pointed by

Sstructures(i) from the Web pages in Sregion(j), construct the inverted-term file by using the I index software, run each query and determine its MZ metric value, record the average MZ metric value for Sstructures(i) and Sregion(j), increase j by 1, if j < 4 then go to filter out the structures pointed, increase i by 1, If i < 4 then go to let the set of the search region, denoted by Sregion; be {r; r with s; r with s with m}.

2.4 WEIGHTED-BASED APPLICATION

According to Huang et al. (2008), it is difficult for users to retrieve information that are special for them, if the search is based on traditional ranking method or the page that the users queried may appear at last of result list. Huang et al. (2008), proposed a new method named Categorization-based ranking algorithm which can help the user to get the target needs from the web pages. The text categorization part straight uses the model created in the classifier construction stage to sort new documents.

Web page sorting, known as web page categorization, might be defined as the task of determining whether a web page belongs to a category or categories. As is shown in Figure 2.4 when the user put query string into Web Server, the Web

Server receive the query and take apart the query string into terms, then search the web pages which are identifying with query terms, authors select the intersection set of web pages, and get out the score of web pages, through getting the score must judge the category that the web pages belong to. If the page belongs to one category the get out the score stands on traditional algorithm, if the web page belongs to not only one category, we must check the categories it belongs to, and get out its synthetically score , then we will get result pages.

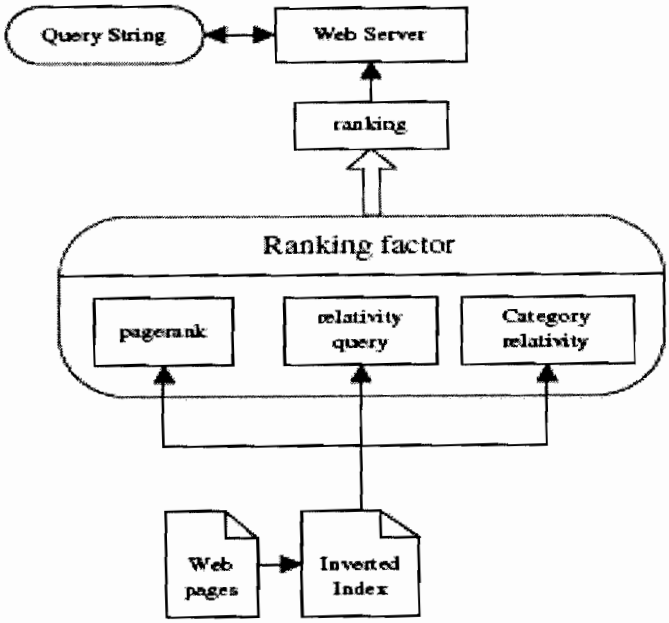


Figure 2.4: The Web Information Retrieval System

(Source: Huang et al. 2008)

belong to ranking algorithm the score of search result is usually low, and cannot appear at the front of the list, the users usually do not has patience to search result page by page.

In categorization-based ranking when we query information during search engine, the result will get out as a list, and the result which is preferences to the user. When we looking to get some information for our preferences, for example, we would like to query car for rent which is the cross of colour and seats, the general method is to select the page which not only belong to colour category but also belong to colour category, and the connection of colour category and seats category are the result.

As described in Lynch et al. (2004), Information Retrieval and Visualization activity is discovering good and new manner particular of organizing and showing the instance or the results that we are queried, describing the data from the database, and helping the user to put the good information that he/she need in his queries. Information Retrieval and Visualization activity have been calculating "similarity ranking" algorithms that give us instance we are looking for in agreement with our similarity to a targets outcome.

Ranking algorithms collected at one time as search result for which related results that users looking for, and the subset of search results provided by the user system. The ranking algorithms work by comparing each subset of search result against the goal, receiving a similarity score for each, then ranking the results agreement with the scores.

This ranking algorithms calculates a similarity score by possessing the two strings as sets of terms A and B, and taking as the score S the percentage of the number of terms in the sets' intersection to the number of terms in their merger.

According to Ranade (1998), if we have a file in the database and we are looking to compute the number of elements that needed it in the file. It is important to supposing that the file elements are stored contiguously in the database, File ranking is interesting because several more substantial parallel algorithms.
In this algorithm, each file element has a numerical value, and the target is to get for each element the summation of the values had in the subfile starting at that element. The problem lessens to file ranking if the numerical value in each element.

As described in Islam et al. (2008), the random walk ranking algorithm on a graph proposed by Brin and Page, has been used in citation analysis, social networks and analysis of the link structure of the web (Blanco & Lioma, 2007). The critical idea carried out by a random walk algorithm is that, when one vertex links to another one, it is basically casting a vote for other vertex (Hassan et al. 2006). The higher the number of votes that are cast for a vertex and the higher is the significance of the vertex.

The good retrieval techniques that done without waste is important in managing the increasing quantity of information existing in the database. Highest in number existent content retrieval techniques depend on indexing keywords. Unluckily, keywords alone cannot suitably get the file content then we will get bad information retrieval. Term weighting is very important for information retrieval system. The target of a term weighting method is to sort the indexing terms by giving them weights similar to how well they make better both recall and precision of the retrieval.

Yan et al. (1996) illustrated an approach for automatically classifying guests of a web site under their access patterns. Authors get the system design that helps them to know the history for quest access patterns to find out common guest access actions.

Guest's access logs are examined to find out clusters of guests that show information like the information that we want. This might outcome in a better understanding of how users visit the site, then it will cause to a better organization of the hypertext documents for finding the way comfortably. Also we can dynamically propose links for them to find the way. In order to explain the general design of a system to carry out these goals, we present an overview as we see in Figure 2.5. The system consists of three parts: a web server capable of maintaining user session information, an offline module responsible for log analysis, and an online module responsible for dynamic link generation

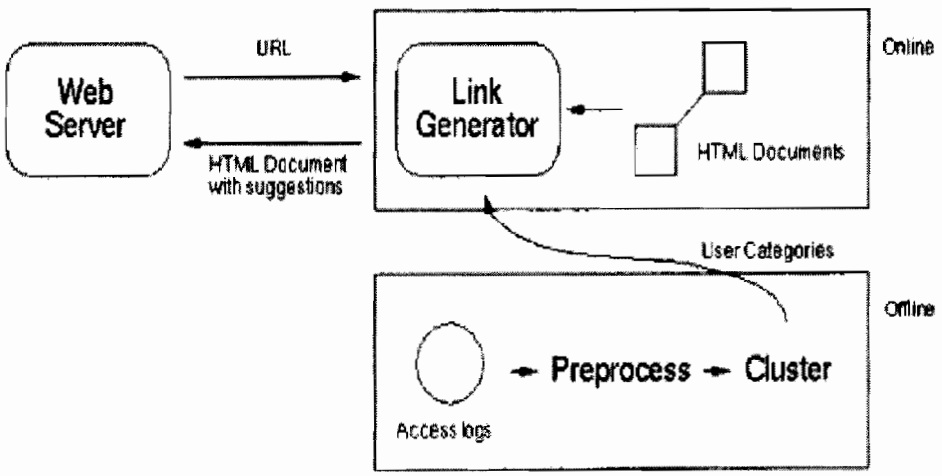


Figure 2.5: overview of system design

(Source: Yan et al., 1996)

First time when the guests use the server, a new session identifier is generated. In the HTML document returned, this identifier is encoded in all URLs applying to objects on the same web site. And when another guest clicks on these encoded URLs, the session identifier is passed back. The online module presents dynamic link generation. When a guest needs and ask about a new page, the module tries to classify his current partial session record against one or more of the categories obtained offline.

In the offline module, the task of the preprocessing step is to exchange the information in user access logs into the vector representation, the preprocessor from time to time take out information from user access logs to make records of user sessions. Every one record generated to one session in the logs. The record registers the access patterns showed by the user in that session. Records are then clustered into categories, with "similar" sessions put into the same category.

According to Bogaschewsky & Hoppe (1993), on the subject of the increase of human being knowledge in the past years and especially the last ten years on sudden increase is evident. seeing as new knowledge is been written down in some way, the knowledge sudden increase is mutually related with an exponential increase of available articles, books, report, story, etc. furthermore, a lot of (mainly textual) non-scientific data, e.g. novels, operator's manuals for household equipment, entertainment literature... etc.

Information retrieval systems (IRS) do not get into explanation the wide range of information needs different users have at various times. Hybrid systems

combining the functionalities of usual information retrieval system and hypertext systems are a promising come near for building user-oriented systems for effective and efficient information retrieval. After the concept of hypertext and the functionality of hypertext systems are outlined, authors talk about how information retrieval and hypertext can be integrated. The specific organization of the information base and the different ways to access the information are analyzed. Furthermore, information retrieval models are outlined and dedicated retrieval models for hypertext are presented.

As described in French et al. (1997), software documentation has many resources that have not been totally taken advantage of; documentation mostly cannot meet the needs of the clients because it is static so gets on a lonely form. Software documentation needs a mix of information retrieval and we will know how the hypertext techniques are the best solution to better increase its utility and best solution for documents fails. And we will see how the prototype system, SLEUTH, now under investigation as way for software documentation management.

As shown in Figure 2.6 we can see the structure that supply consumers with software documentation that has a much richer structure than usually the case, the author uses the listed hypertext information to set at a starting position the cross-reference filters.

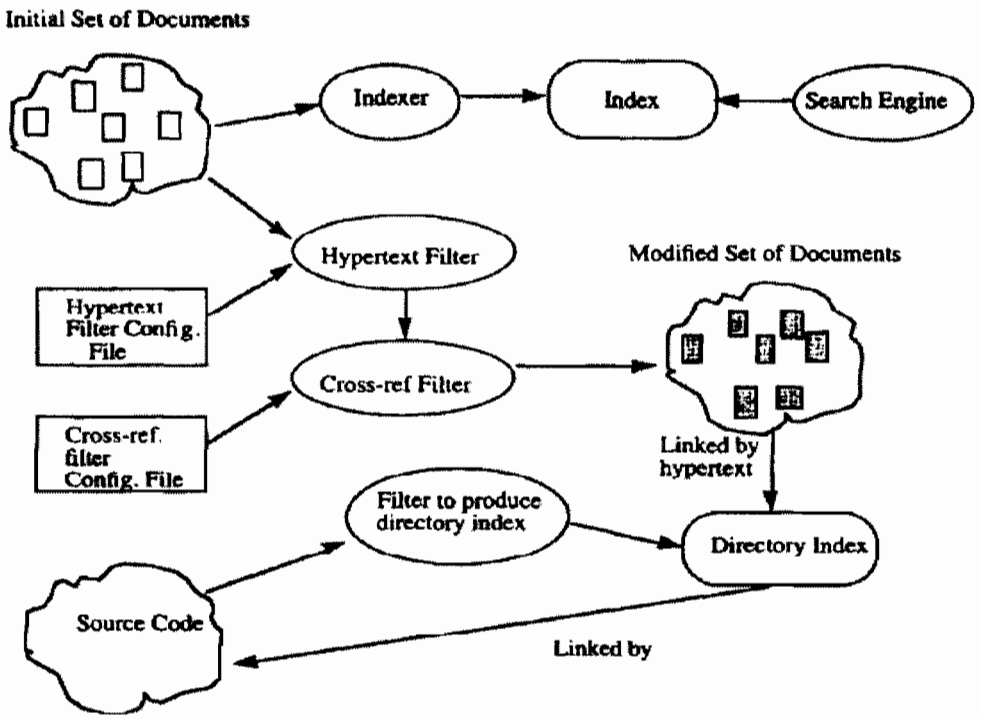


Figure 2.6: SLEUTH authoring architecture

(Source: French, et al. 1997)

Select concept alternatives. Decision matrix method is a method that lets decision makers to methodically make out the strength of relationships between sets of information. Decision matrix method is looking at a numbers of factors and getting each relative importance. However, decision matrix method is a way for alternative selection using a scoring matrix.

Decision matrix method, teams selects a list of weighted criteria and then evaluates each alternative against the old criteria. Then need to hand over a relative weight to each criterion. This handing over must be done by team consensus. In addition, each user can hand over weights by them self; the alternatives are gained concerning to them criteria (Salmeron,& Smarandach, 2007).

2.5 CONCEPT BASED TERM WEIGHTING

According to Zakos, J., & Verma, B. (2005), The aim of CBW is, given a query term q and the WordNet ontology as inputs, to calculate the term importance for q by interpreting the conceptual information related to q in WordNet and giving a resultant value in the range [0, 1]. A value of 0 or close to 0 indicates a term with low importance and a value of 1 or close to 1 indicates high importance. The technique is based on the notion that the more general or vague a term is, the less important it should be. Also, the more specific and precise a term is, the more important it should be Figure 2.7 represent the CBW. CBW exploits and interprets 4 types of conceptual information in WordNet for a given term: 1) number of senses 2) number of synonyms 3) level number 4) number of children. We propose that term generality vs. specificity can be derived from these 4 types of conceptual information and that term importance can be calculated as a consequence. and divided into three steps as the following: 1 extract conceptual CTM, 2 transform the integer values in the CTM into weighted values, 3 fuse the weighted values in the CTM to give a single and final CBW importance values for term Q.

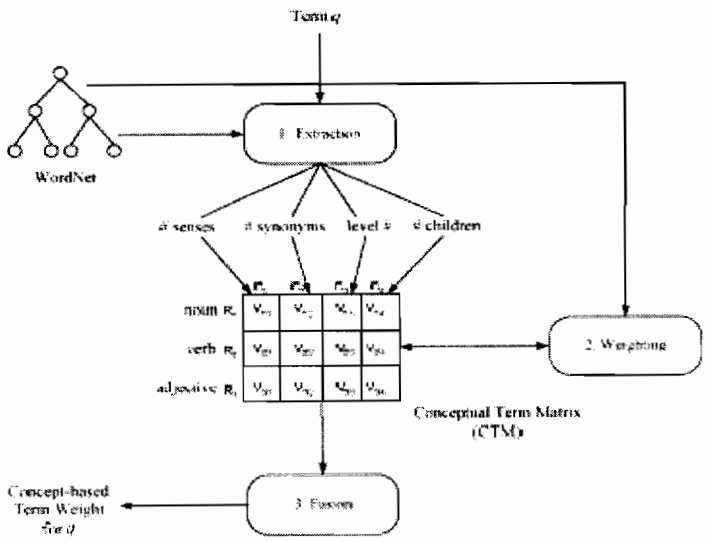


Figure 2.7 Overview of CWB

This first step in the concept base term weighting is to capture the term form the user to calculate the importance. CBW have matrix m x n conceptual term to holds the m conceptual information types for N section represent ontology for each term Q. and in the Extraction task is to extraction to obtain an integer value for each of the 4 types of the conceptual related to term Q, and use the extraction algorithm to calculate integer value for each conceptual, and the next step is Weighting, in weighting we weight the integer values in the CTM given by extraction to obtain the final weight values for each term Q, and the last step is fuse, The aim of fusion is to combine all the elements in the matrix to obtain one single value that is indicative of term importance. While different fusion methods may be used, the fusion procedure always occurs by first fusing columns and then the row:

1. Fuse each column of the CTM separately using a fusion function f1(C). This gives a resultant row R where each element in the row represents a fused column.

R = { f1(C1), f1(C2), f1(C3), f1(C4) }

2. Fuse the row R generated in Step 1 with a fusion function f2(R) to give the CBW term importance:

CBWq = f2(R)

The 3 standard fusing functions for both row and columns are AVG, MIN and MAX. AVG averages the elements in the vector to give an average value. MIN selects the smallest element value and MAX selects the largest element value. A more advanced fusion function that can be applied is weighted average:

$$CBW_q = \frac{\sum_m R_m \times W_m}{\sum_m W_m}$$

Where W is a set of weights with each element being a value in the range [0, 1]. This fusion function has the ability to weight the importance of the different conceptual information types.

2.6 SUMMARY

After we understood the user needed and studied the information retrieval and search engine and we found more algorithms that were an adopted in information retrieval to get user satisfaction, I suggest the concept base term weighting as algorithm to apply in second hand car to improve the information retrieval in this web site and give some facilities to user during his using the second hand car website, Zakos, J., & Verma, B. (2005), The best CBW result performs comparably against the traditional term importance measure IDF. CBW improves precision for many queries and could potentially outperform IDF with further investigations and advancement of the technique. Since initial experiments utilizing weighted average for row fusion have given the best CBW results, we plan to investigate more advanced and weighting

fusion methods in an attempt to optimize CBW to improve its effectiveness. So the reason of using concept base term weighting algorithm is because more previous researches approved that this algorithm is better than other algorithms that used in the same area of this study.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 INTRODUCTION

In the previous chapter, we presented discussions on the literature related to information retrieval, search engine and Concept based term weighting (CBW). This chapter elaborates the research methodology which is adapted to this study. The methodology is an agreeable method, excellently chosen, described and accepted among the experts in Information System Research Design (Vaishnavi & Kuechler, 2007). Overview of the methodology is briefly discussed in Section 3.2.

3.2 METHODOLOGY

In this work we use one of the most popular realizations of the iterative approach for object oriented development which is the Rational Unified Process. Sines its object oriented it will use Identifying Actors and Use Cases, Use Case Modelling and Analysis, Object Modelling, UML to represent the objects that are used in the project,

these are a great starting point. Language based on Data-Flow Diagrams, Structure Charts, and other elements), then become ready to the processing phase using the Rational Rose technique. (Bennett *et al* .2006).

The Rational Unified Process consists of cycles that may repeat over the long-term life of a system. A cycle consists of four phases: Inception, Elaboration, Construction and Transition. Each cycle is concluded with a release, there are also releases within a cycle. Let's briefly review the four phases in a cycle. (Bennett *et al* .2006).

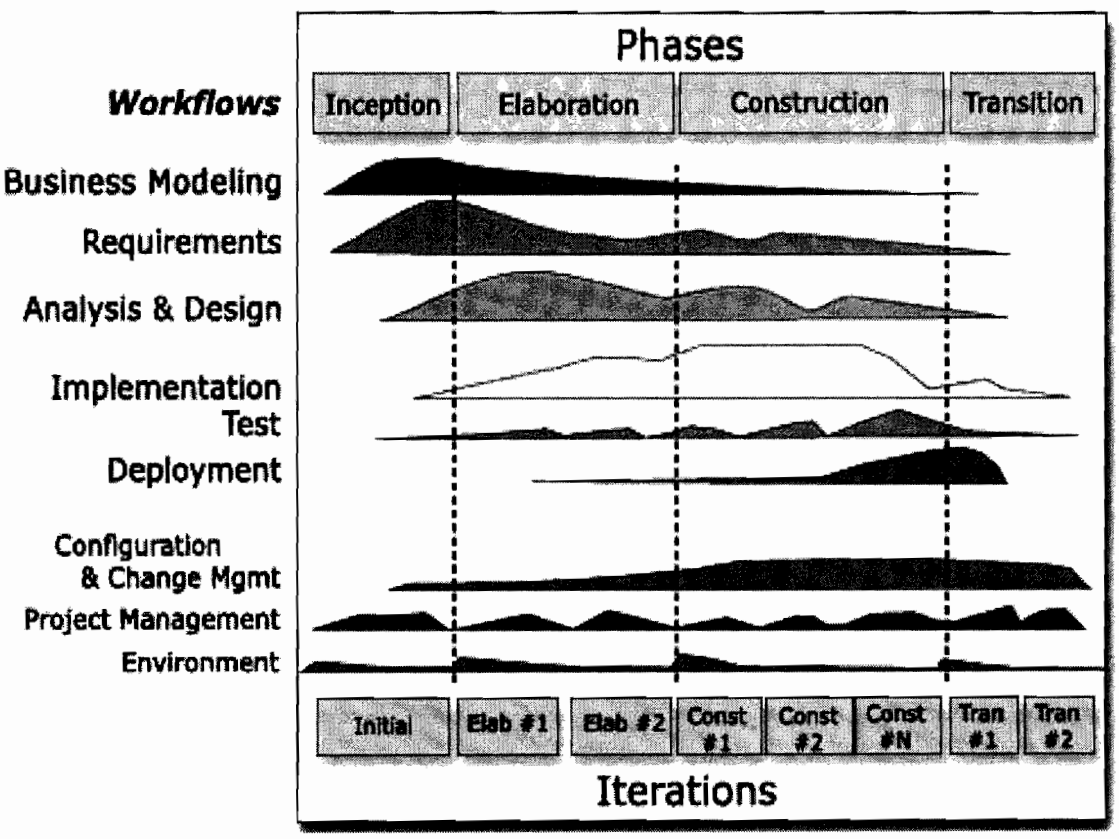


Figure 3.1: Rational Unified Process. (Bennett et al .2006).

3.2.1 Inception Phase

During the inception phase the core idea is developed into a product vision. In this phase, we review and confirm our understanding of the core business drivers. We want to understand the business case for why the project should be attempted. (Bennett. R *et al* .2006).The inception phase establishes the product feasibility and delimits the project scope.

- In this stage we got to know the gap between the user requirement and available products then define the problems statements.
- Based on our observation there are some limitations in the existing online second hand car application such as existing online do not provide customers with the opportunity to include their requirement in buy or sale a vehicle. Such a facility would allow a customer to prioritize his/her preferences.
- The main problems are mentioned as bellow:
  - A limited number of personal preferences are provided by the company. For example www.mudah.my does not provide the car info.
  - Customer would not prioritize his preferences.
  - The search for relevant vehicle to be buying is based on a single database. Hence, a user has to visit other websites if he/she requires other options.

Based on these problems, we see a need to design and develop a new approach in identifying relevant information. The output of this phase is a tentative design that includes information on online second hand car trading from two companies in

Malaysia, information retrieval used by these companies and formulation suitable algorithm Concept based term weighting to achieve the desired result.

3.2.2 Elaboration Phase

During the elaboration phase the majority of the Use Cases are specified in detail and the system architecture is designed. This phase focuses on the "Do-Ability" of the project. We identify significant risks and prepare a schedule, staff and cost profile for the entire project. In this phase we will build a logical model that will represent the system. A use case describes how the user will interact with the system to perform some activate. (Bennett. *et al* .2006).

i. List of Requirements

In general, list of requirements are partitioned into functional requirements and non-functional requirements. Functional requirements are associated with specific functions, tasks or behaviours the system must support, while non-functional requirements are constraints on various attributes of these functions or tasks. The functional requirements address the quality characteristic of functionality while the other quality characteristics are concerned with various kinds of non-functional requirements. Because non-functional requirements tend to be stated in terms of constraints on the results of tasks which are given as

functional requirements (e.g., constraints on the speed or efficiency of a given task), a task-based functional requirements statement is a useful skeleton upon which to construct a complete requirements statement. That is the approach taken in this work. It can be helpful to think of non-functional requirements as adverbially related to tasks or functional requirements: how fast, how efficiently, how safely, etc., is a particular task carried out by the second hand car web site, Table 3.1 represents the functional requirements and table 3.2 represents the non-functional requirement of the system. In the priority column, the following short hands are used:

- M – mandatory requirements (something the system must do)
- D – desirable requirements (something the system preferably should do)
- O – optional requirements (something the system may do)

Table3.1: Functional Requirement

No.	Requirement ID	Requirement Description	Priority
		VISITOR REQIREMENTS	
1.	**SHC _01**	**Explore System**	
2.	SHC _01_01	View General information about all company	M
3.	SHC _01_02	View register requirement.	M
4.	SHC _01_03	View website features.	M

5.	SHC _01_04	View all new cars available to sale.	M
6.	SHC _01_0	The visitor can print all the General Information	D
11.	**SHC_02**	**Apply to sale car.**	D
12.	SHC _02_01	Visitor can apply to sale car by click SALE button	D
13.	SHC _02_02	Insert car information into the form's field.	O
14.	SHC _02_03	Visitor can reset all car information by click RESET.	D
15.	SHC _02_04	Visitor can click CANCEL button to return back.	O
16.	SHC _02_05	Displays error message if incomplete car information.	
17.	SHC _02_06	Visitor can click SUBMIT button to sent the application.	
18.	**SHC_03**	**Apply to buy Car.**	
19.		Visitor can apply to sale car by click BUY button	
20.	SHC_03_01	Insert car information into the form's field.	M
21.	SHC_03_02	Visitor can reset all car information by click RESET	D
22.	SHC_03_03	Visitor can click CANCEL button to return back.	M
23	SHC_03_04	Displays error message if incomplete car information.	
24.	SHC_03_05	Visitor click SEARCH button to retrieve the information.	

LOGIN TO THE SYSTEM			
	SHC_04	**Login**	
25.	SHC_04_01	User must key-in user ID and password.	
26.	SHC_04_02	Use press LOGIN button.	
27.	SHC_04_03	User can click CANCEL button to return back.	
28.	SHC_04_04	System will verify the Login information	
29.	SHC_04_05	Return missing Login information click FORGET button.	D
REGISTRATION			
30.	**SHC_05**	Insert personal information into the form's field.	M
32.	SHC_05_02	Visitor can reset all personal information by click RESET	M
33.		Visitor can click CANCEL button to return back.	M
34.	SHC_05_02	Displays error message if incomplete user information..	M
35.	SHC_05_03	Visitor can click SUBMIT button to sent the application.	M

Table 3.2: Non_ Functional Requirement

No.	Requirement ID	Requirement Description	Priority
1.	**SHC _04**	**Reliability issues**	
2.	SHC _04_01	For a single user, the system should crash no more than once per 10 hours.	M
3.	SHC _04_02	If the systems crash, it should behave perfectly normal when reloaded again.	M
4	**SHC _05**	**Usability**	
5	SHC _05_01	Improve the system interface.	M
6	SHC _05_01_01	Easy to use and don't need training	M
7	SHC _05_02	Categorize depended on the subject	M
8	**SHC _06**	**Efficiency**	
9	SHC _06_01	Performance	D
10	SHC _06_01_02	Ability to search and retrieve data	D
11.	SHC _06_02	Space	D
12	SHC _06_02_01	Support the system with a large data bases	D

The table of information needed in the functional requirement are:

1. General Information: (company name, address. Phone numbers, Company features, register requirements, fees).
2. personal Information :(name, national number, address, phone number, mobile number and email).
3. Car Information: (kind, model, color, year, airbag system, Air-condition, car engine, gear type, body).
4. Login Information: User name and Password.

ii. Requirement Modeling

A Unified Modeling Language (UML) is used to visualize and document the software system design. UML uses object-oriented design concepts which can be uses to describe the business process and requirement generally. The objective of UML is to provide a common vocabulary of object-oriented terms and diagramming techniques that is rich enough to model any systems development project from analysis through implementation (Bennett. R *et al* .2006).. Here, a use case and sequence diagram are drawn to represent the activities involves in the system.

iii. Use Case

A use case is a sequence of action that allows the analyst to model the interaction of information system base on its environment. A use case is drawn usually to reflect when an actor uses a system to complete a certain task or process. The actor may be ranging from a person (or persons), a computer device, computer system or even and organization. Figure 3.2 represents the general behavior of the second hand car trading. We represent the use case specification of five main function requirement of the second hand car trading on the appendix A.

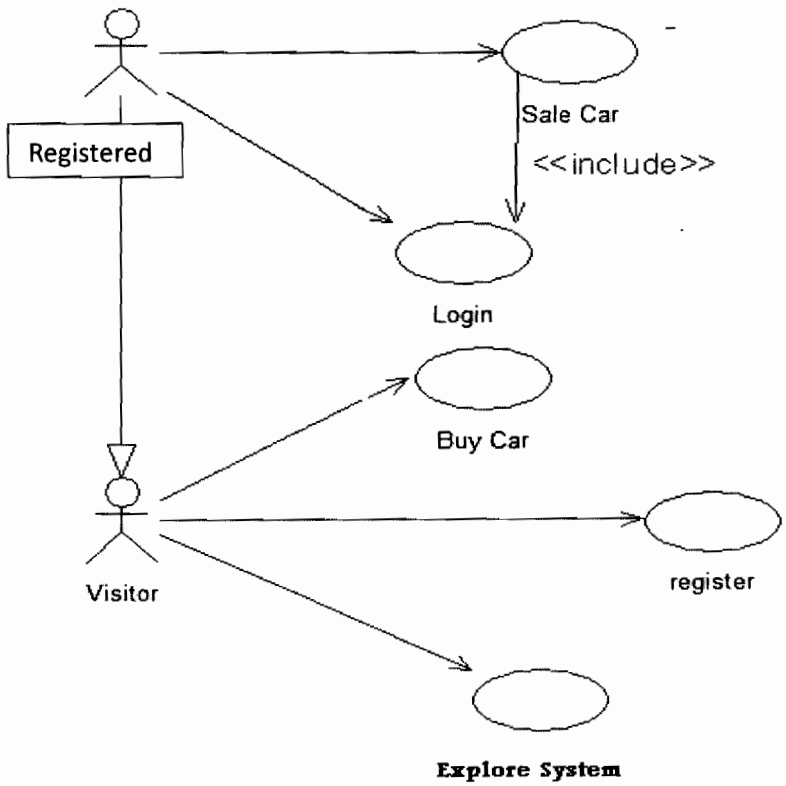


Figure 3.2: USE CASE DIAGRAM

iv. Sequence and Collaboration Diagram

A system sequence diagram is drawn to show the interaction between a actor and the system, for one use case scenario. According to Dennis *et al* (2005), sequence diagram permits and analyst to portray the dynamic interaction among object in an information system and are the most common kind of interaction diagram used in object- oriented- modeling. Figure3.3, Figure3.4 and Figure3.5 represent the sequence diagram base on the use case of three main function requirements for the second hand car trading. for more details about sequence and appendix A.

Collaboration is a set of object and relationships in a particular context. Collaboration diagrams are static construct to show objects and messages involved in accomplishing a purpose or a set of purposes. Since time is not shown as separate dimension by sequence numbers, (Dennis.*et al*, 2005). Appendix A represents collaboration in more details.

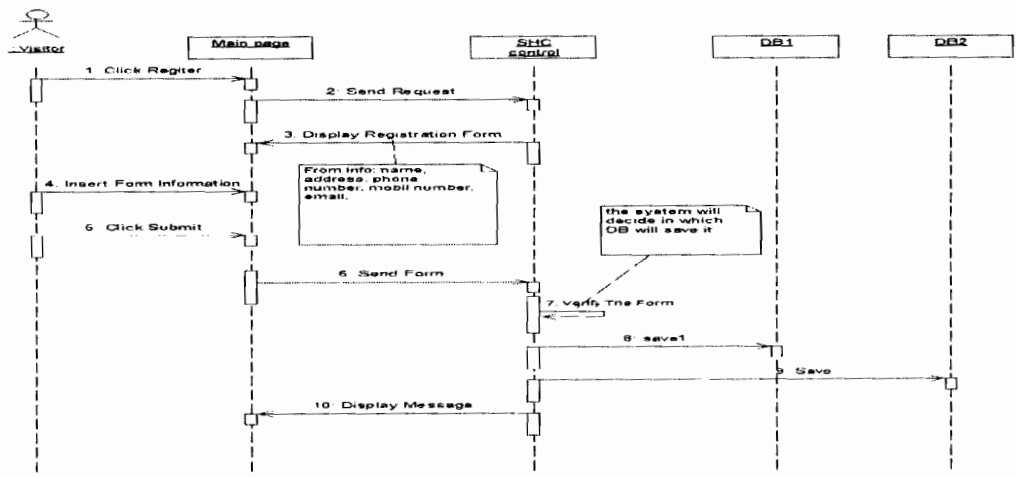


Figure 3.3: Sequence Diagram of the Registration System

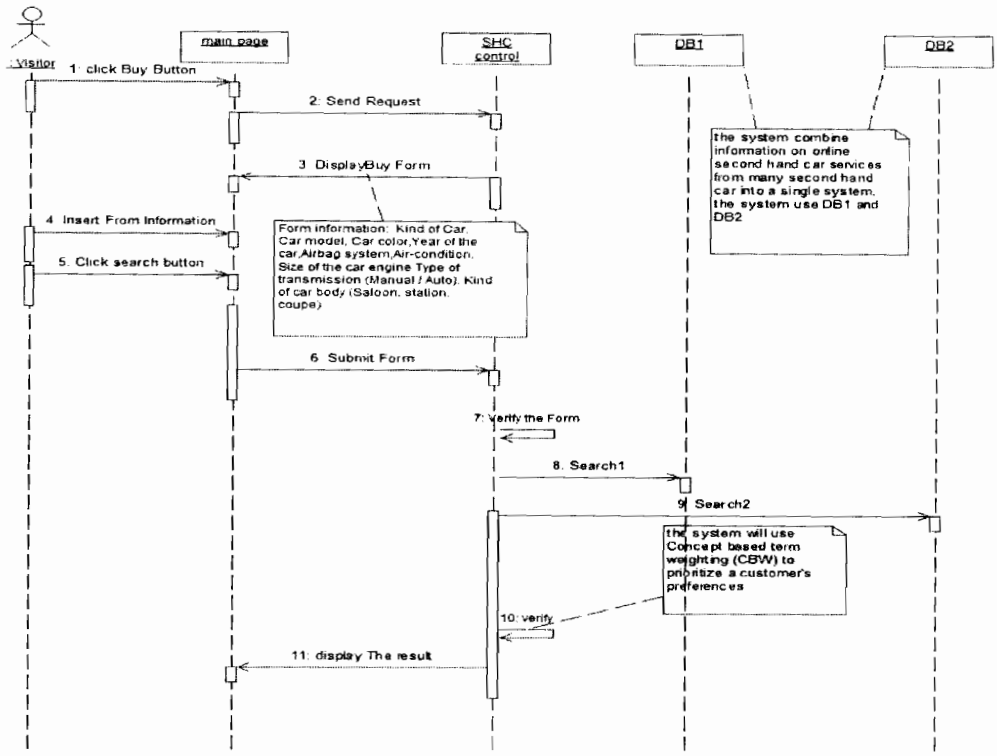


Figure 3.4: Sequence Diagram of the Buying System

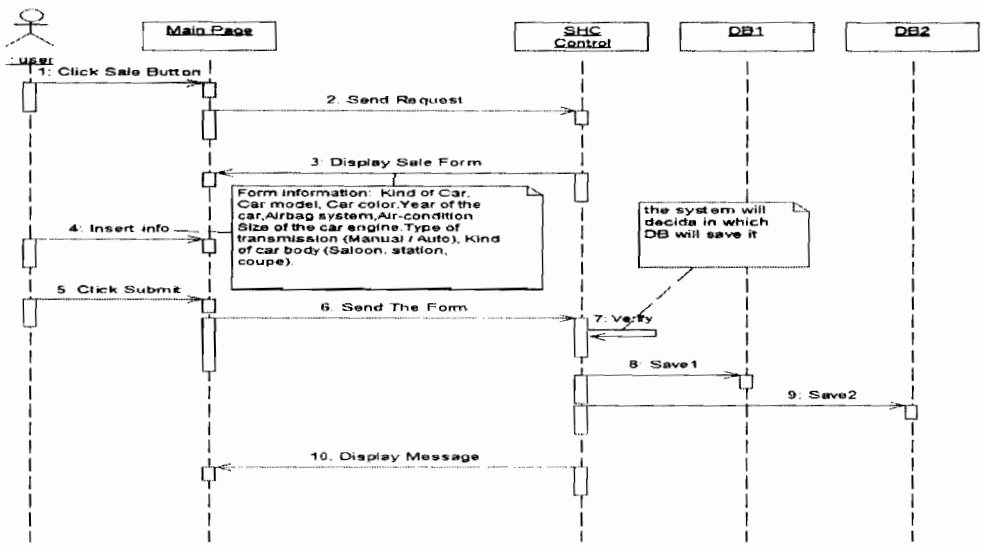


Figure 3.5: Sequence Diagram of the Sale System

v. Activity Diagram

An activity diagram is special from for state diagram where the states represent performance of operations and the transitions are triggered by the completion of the operations (Dennis *et al*, 2005). Figure 3.6 represent the activity diagram for the system.

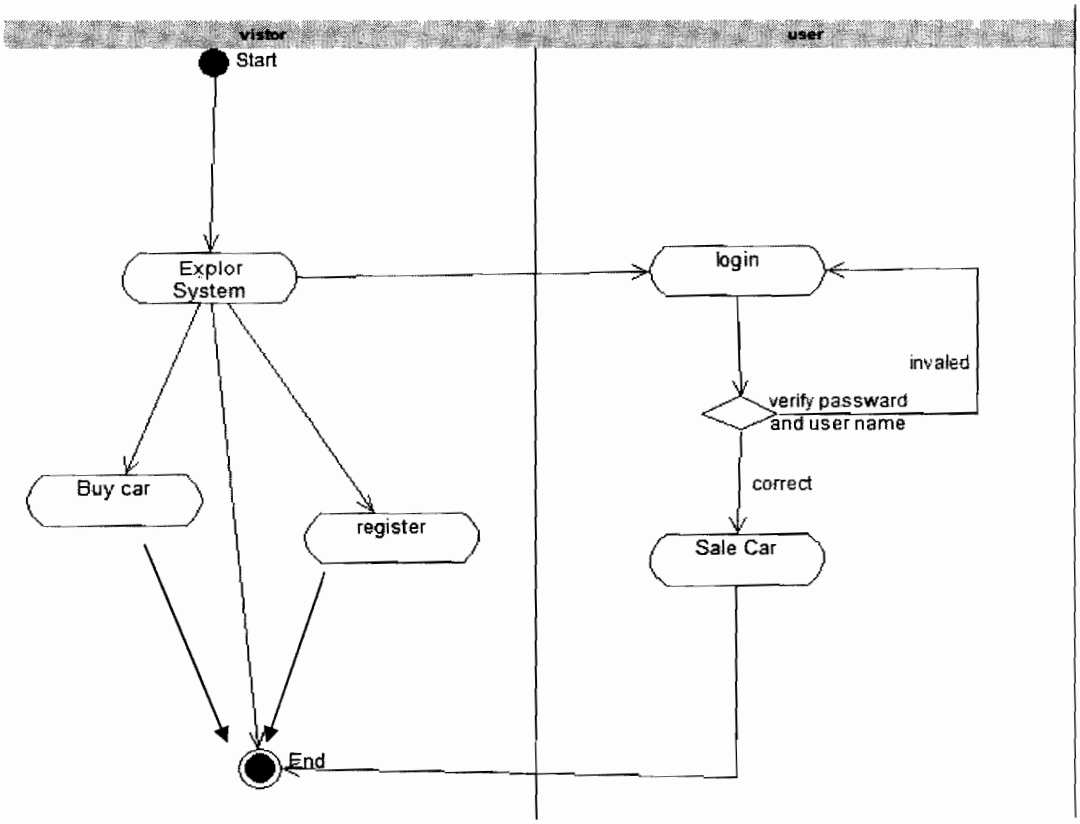


Figure 3.6: Activity Diagram

vi. CLASS DIAGRAM

Class diagram describe the static structure to a part of a system. The diagram does not only show classes connected by static relationship (associations), but also packages, interfaces, objects and links etc. Specification of behavior of a class' objects is not supported. This is regarded as part of property modeling and described in terms of functional roles by means MSCs. During design the behavior of classes will correspond to behavior of SDL processes and may then be defined by process graphs. Figure 3.7represent the class diagram to second hand car trading.

.

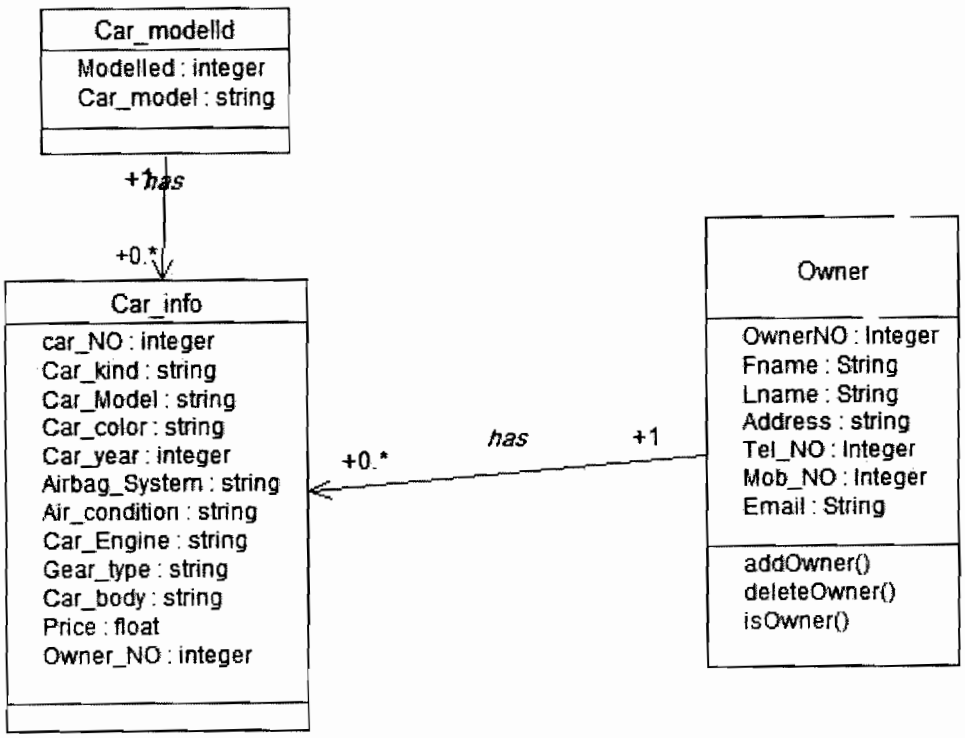


Figure 3.7: Class Diagram

6.2.3 Construction Phase

During the construction phase the product is moved from the architectural baseline to a system complete enough to transition to the user community. The architectural baseline grows to become the completed system as the design is refined into codes. The design is implemented in this phase as follows.

I. Database

The Web-Based second hand car trading prototype is developed. We use the SQL server 2005 to create the necessary tables which store information on the second hand car trading application. In this project we created two second-hand car databases.

Each database consists of three tables' car information table, owner table, model table.

A. Car information table

Car information is an entity which has the following information as attributes:

- **Owner id:** user can use this owner id to owner information e.g. phone number and address
- **Car kind:** User has choice to choose the kind of car that he/she prefer to buy for different car available e.g. proton, BMW, Toyota, and nisson.
- **Car model**: User has a choice to choose in between different cars available e.g. Saga, Waga, Wira, Cenetra, and Perdana.

• **Engine Capacity**: The user can give the limitation of engine capacity. For example, user can ask for the cars having capacity 1600.

• **Colour**: The user can input the colour choice of the car that he/she prefer.

• **Gear Type**: The user can choose between manual and automatic car.

• **Air Bags**: The user can ask for the cars only with the air bag available for security reasons.

• **Air Condition**: The user can ask for air conditioned or non-air conditioned car.

• **Kind of body**: User has choice to choose the kind of body that he/she prefer e.g. Saloon, station, and coupe.

• **Price**: The user can limit the price of the cars to be searched.

B. Owner Table

Owner is an entity which has the following information as attributes: owner_no Fname, Lanme, phone number, mobile number Email and the address.

Table 3.3: Owner Information

	Owner_NO	Fname	Lname	Address	Tel_NO	Mob_NO	Email	password
▶	3	saed	tahat	kedan	233445	23452346	saed	3
	5	omar	omar	sa	1111	111	da	5
	22	bander	ali	penang	123456789	12345678	asdfasdg	22
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

C. Model Table

The model is an entity which has the following information as: model ID and car model.

Table 3.4: Modeled

	modelId	Carmodel
▶	1	proton
	2	Bmw
	3	mercedes
	4	AUDI
	5	nissan
	6	Toyota
	7	Mazda
	8	Suparu
	9	Honda
*	*NULL*	*NULL*

To the complete information about the database, DB1 and DB2 can be referred in Appendix E.

II. System Prototype

In order evaluate the effectiveness of concept based term weighted (CBW) in information retrieval; we develop two versions (A, B) of the system. in(version A) the search is performed based on hypertext-matching while (version B) uses the concept based weighted approach.

A. Version (A)

In version A the search is performed based on hypertext-matching, Version A works in the following way. For example, if we want to find the Saga car having a black colour and manual gear type from two second hand car companies then the results.

Figure 3.8 shows how the Hypertext-matching analysis algorithm works. User sends a query to search of the following:

1. Car model = Saga
2. Colour = black
3. Gear type = manual

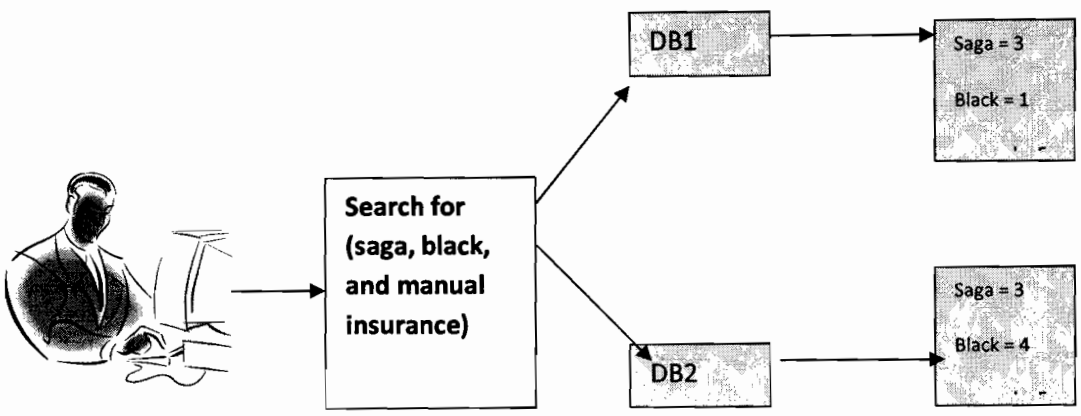


Figure 3.8: Hypertext-matching Analysis Algorithm.

The system will first search the word "saga" the two databases, then it will search for word "black" and then it will search for "manual". As shown in the Figure 3.8 in DB1 there are three Saga cars, one black car and five cars that have manual gear. This total of folded documents is 9, while in DB2 the number of relevant documents is 12. Results from the database that has the most search result (i.e. DB2) will be presented on the top of the retrieval list.

Following is the screen shot (Figure3.9) of the search screen that appeared when the user choose buy car link.

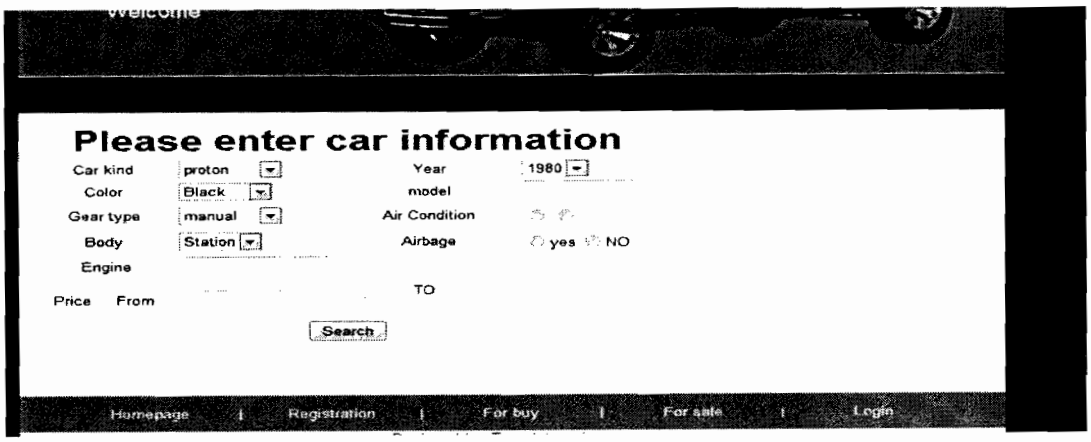


Figure 3.9: Search Page of The System (version A)

B. Version (B)

In version B we include additional information indicated by a user. The additional information is the priority search. A user can select a priority value to each of the

nine attributes presented in the search form, the priority number is a number between one and ten, number one is the highest priority while the number ten is the lowest priority. The concept base term weighting calculates the priority for each car and retrieve the car that have highest priority For example, a user can give first priority to car model, second priority to car insurance and third priority to car colour and so on. Figure 3.10 is screen shot of the search page

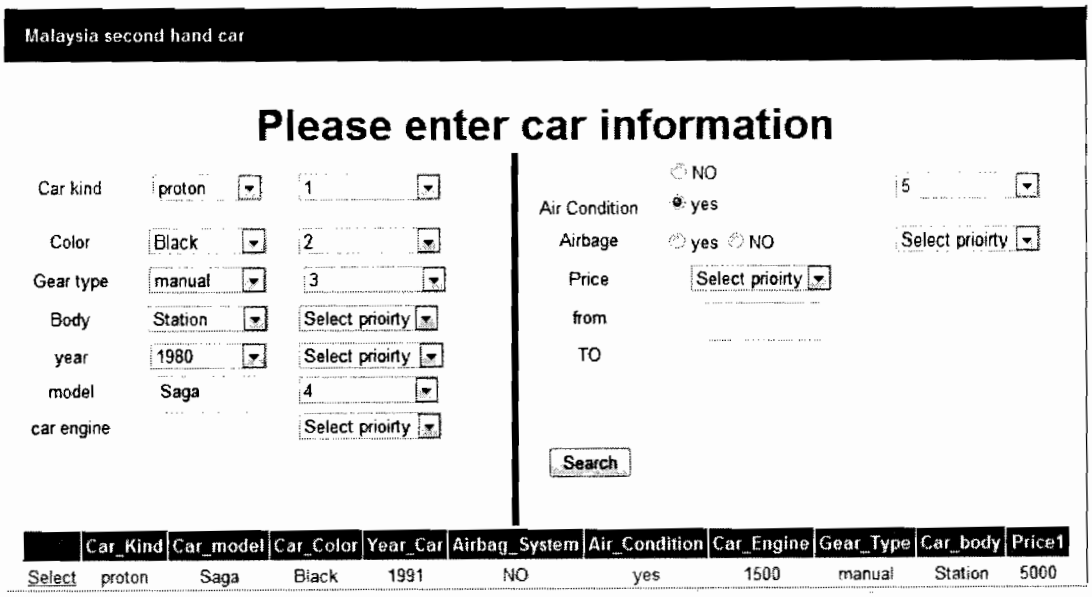


Figure 3.10: Search Page of the System (Version B)

Version B use Concept base term weighting algorithm, the aim of CBW is, given a query term q as inputs, to calculate the term importance for q by calculates the conceptual priority number related to q and giving a resultant value in the range [0, 1]. A value of 0 or close to 0 indicates a term with low importance and a value of 1 or close to

1 indicates high importance. As our project is about to give the best matching result of the query which the customer has asked, this algorithm is most related to our study.

Figure 3.12 shows how the concept base term weighting approach works. If the user wants to find a car with the following details:

1. Car model: Saga
2. Colour: Black
3. Gear type: manual
4. Air condition

He/she then needs to prioritize his/her search requirements. For example, he/she may prioritize these requirements as 1, 2, 3 and 4 respectively.

The system will search in DB1 and DB2 for the required car. And the concept base term weighting will calculate the priority for each car; the result will be save in a new database table that's later is sorted in a descending order.

In DB1, when the system searches the first record, we will identify the record as 1204 meaning that the car fulfil all of the required information except for the third attributes (identify by 0). If the system does not find the required information then it will assign 0 to the particular attribute of the record. From Figure 3.11 we can see that there is one record that fulfils all of the required details (first record in DB1).

Results from the two databases will be combined into a single Table that later will be sorted descending. For example, system will display 1230 (higher number) first and then 1204 (lower number) second.

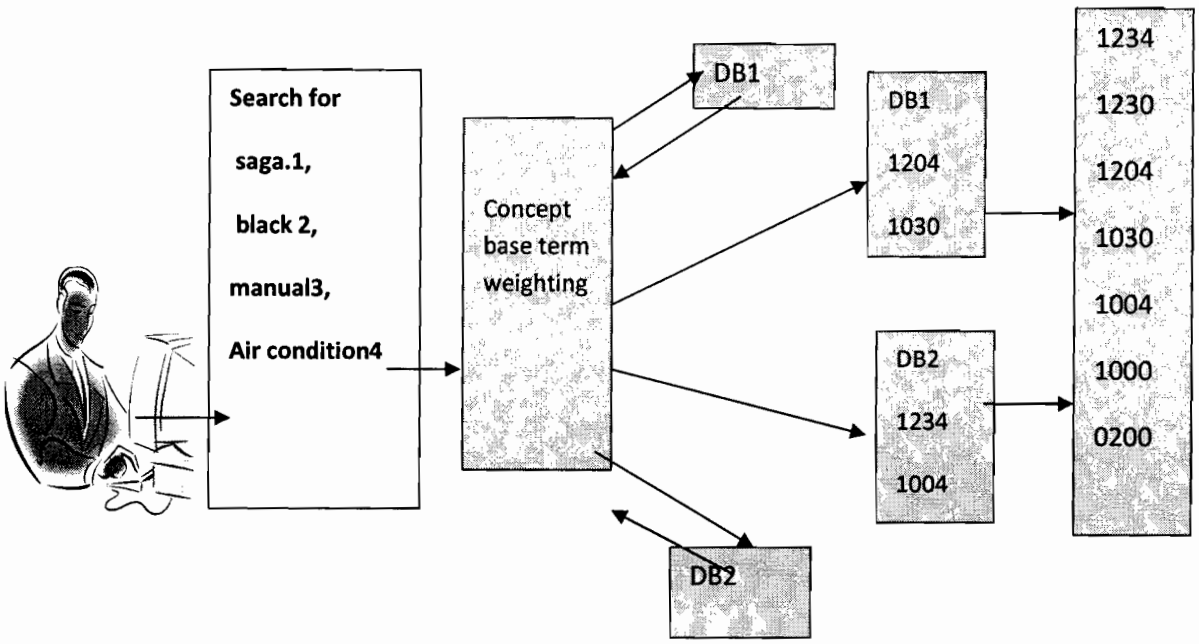


Figure 3.11: Preferences First Search Algorithm.

C. System Development

For the system development, ASP dot net were used for the Web pages, with the web developer 2005 framework, which is capable of providing the virtual web server necessary to run the programs.SQL server was used for the Database to store and retrieve all information, figure3.12, figure3.13, figure3.14, figure3.15 show the main page of the second hand car trading web site.

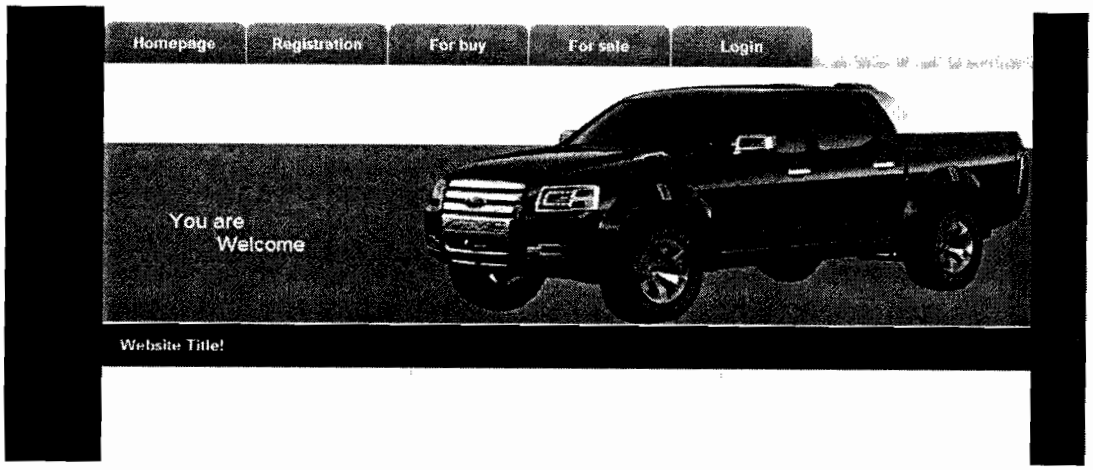


Figure 3.12: Main Page

Description: The main page has four links; registration, for buy, for sale and login, the User will choose that link he/she wants.

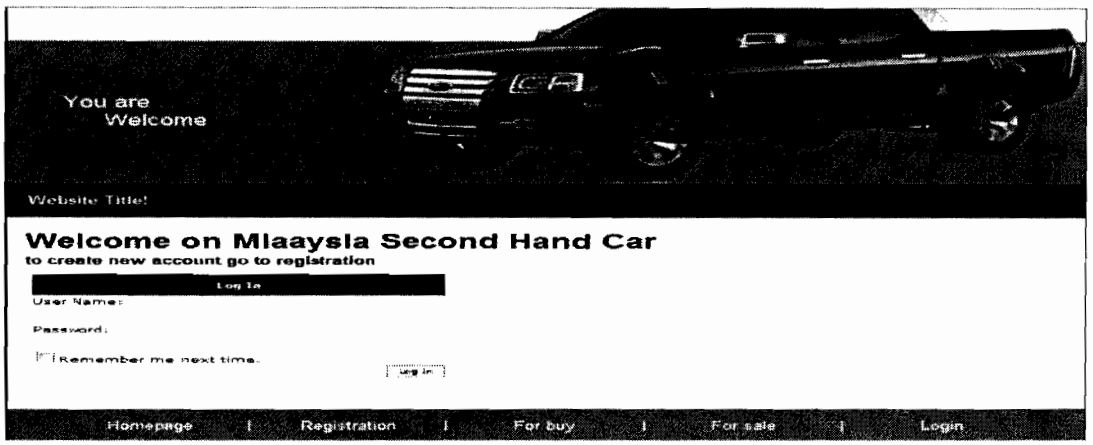


Figure3.13: Login Page

Description: Login page each user have user name password, if u user want to sale has car most enter has username and password.

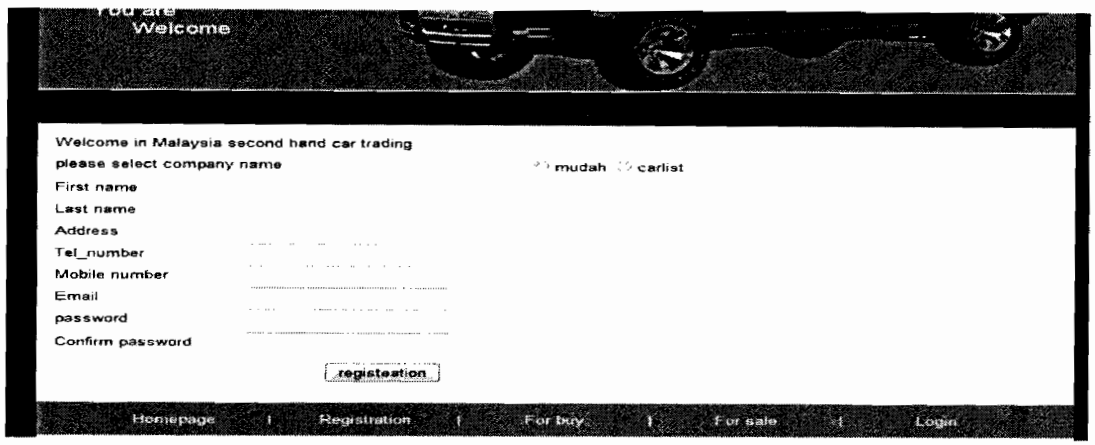


Figure3.14: Registration page

Description: Users are required to fill in the required information such as first name, last name, and address, etc, to register as new user.

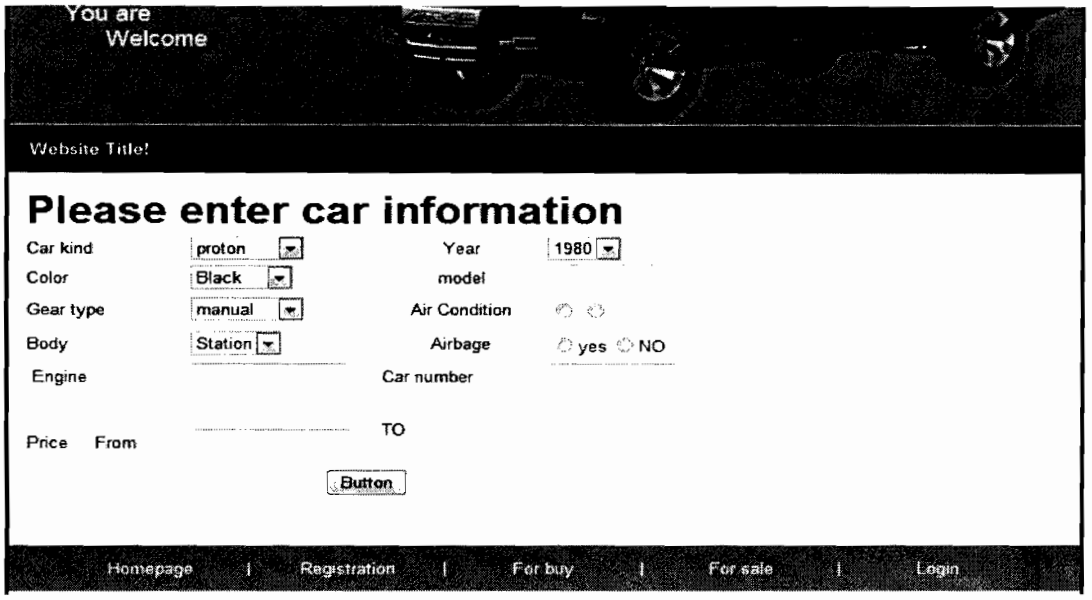


Figure3.15: sale page.

Description: For sale page allow the user to enter the car in formation that he/ she need to sale it.

3.2.4 Transition phase

In the transition phase the goal is to ensure that the requirement have been met to the satisfaction of the stakeholder. This phase is often initiated with a beta release of the application. Other activities include site preparation, manual completion, and defect identification and correction. The transition phase end with a post-mortem devoted to learning and recording lessons for future cycles. (Bennett. R *et al* .2006).

According to Williams (2009) the usability testing in a system of software system development projects can be considered as the most important part. Usability effectiveness, efficiency and specific users can achieve specified goals in a certain environment to create satisfaction

The System testing is the final stage in the system development, after the system has been developed; it was tested by running the system, Dennis (2002) system testing is to minimize system from bugs and errors but not to fully eliminate it. The user evaluation of the prototype was conducted on thirty respondents; each of them was given brief explanation regarding the usage and the user interface of the prototype.

To ensure the requirement have been met to the satisfaction of the user, thirty students from UUM were asked to examine the prototype and provide their feedback. We put the results that got from students in SPSS to evaluate the results. These results will explain in Chapter 4.

3.3 Summary

In this chapter, the methodology which is suggested for this study is presented. A generalized sequence of the several steps for building and deploying traditional and enterprise applications is discussed. There are two Databases created for the power of the experimental the search, The Database store information of the cars. The version B includes additional information indicated by a user. The additional information is the priority search.

CHAPTER 4

EVALUATION AND RESULT ANALYSIS

4.1 INTRODUCATION

This chapter presents the results of the system test and evaluation uses usability testing based on the standard tests and collected by the researcher. Testing with potential users can obtain as efficient feedback as possible in a short time frame and with the available resources. It is also irrelevant to ask people in a focus group to predict whether they would like something they have not tried, so the only way to get valid data is to let users experience the technology before opinions are sought (Nielson, 2000).

According to Williams (2009) the usability testing in a system of software system development projects can be considered as the most important part. Usability effectiveness, efficiency and specific users can achieve specified goals in a certain environment to create satisfaction.

The questionnaire was adapted from (Davis, 1989), it covers two dimensions: Usefulness, and Ease of Use of the prototype. The questionnaire (Appendix A) consists of two sections: General information and System aspects.

The system evaluation measures the system usability that adopted for the evaluation of the weighted-based approach in representing user preferences in information.

4.2 Researcher Evaluation

To evaluate our system as presented in chapter two, there are two versions of the system. Version A is build based on hypertext matching while version B applies the Concept based term weighting (CBW).Each of version A and B have the same environment like the database and the car information, but the different is version B adopt concept base term weighting to improve user preference, In the new version B user can select the priority for the car attribute that he/she prefer, to evaluate our system before the user to ensure from the system efficiency we will take example(proton, Gen2, silver, air condition, airbag system and manual).firstly we make search on the version A and version B using this example and we will compare between the results. Figure 4.1 represent the search in version A with the retrieval information while Figure 4.2 represent version B with the ten best results.

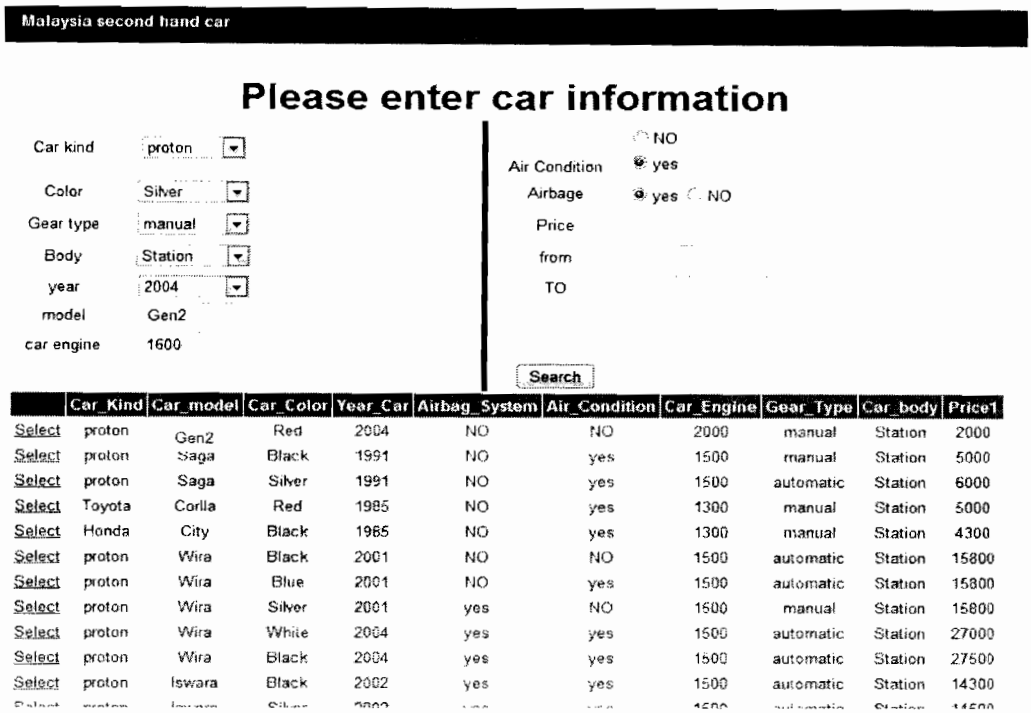


Figure4.1: Search Page Version A

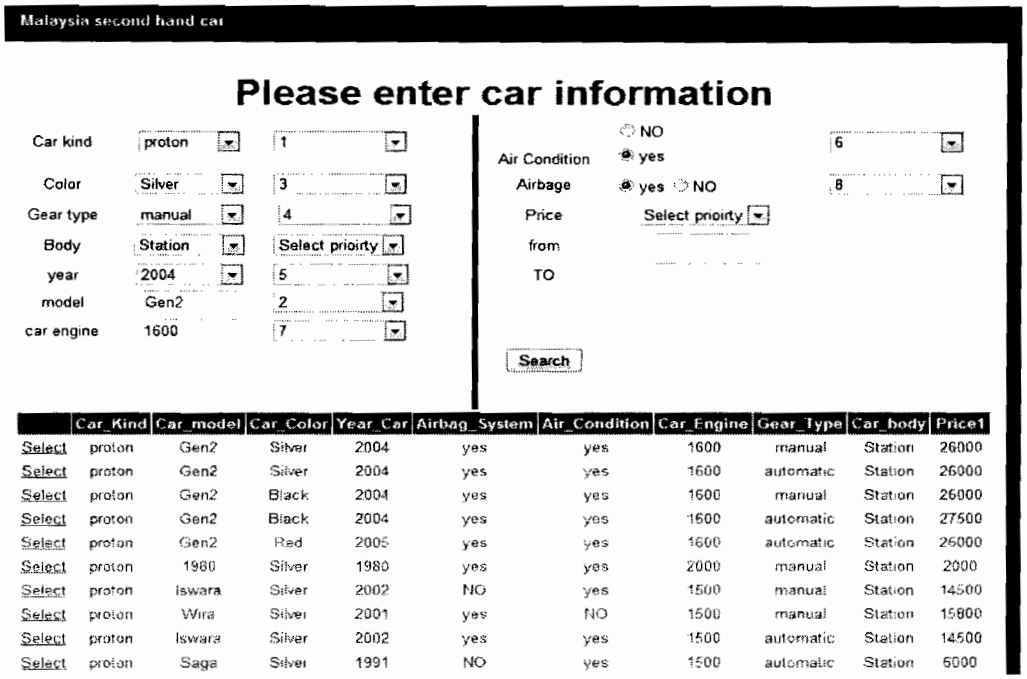


Figure4.2: Search Page Version B

As we see in figure 4.1 and figure 4.2 same user requirement but different results. The different is very clear. At version B we see the first result is (proton, Gen2, Silver, manual) and this car matches the user preference, and the second result also same but it not manual it automatic and the next car is also (proton, Gen2, black manual), we got this result cause the user select the high priority number one for car kind (proton), second priority for car model (Gen2) and third priority for the colour (Silver) est. Then the concept base term weighting calculate the priorities for each car in the database one and the database two and retrieve the best ten result as we saw on figure 4.2. While at version A we see the Hypertext-matching analysis algorithm and the result that this algorithm gets. The system Based on the result presented this exam, we concluded that the respondents felt version B is better than version A. This shows that by applying a Concept based term weighting (CBW), relevancy of the retrieved results is improved and at the same time fulfilled user preferences.

4.3 User Evaluation

The system evaluation measures the system usability that is adopted for the evaluation of the concept-base term weighting approach in representing user preferences in information. The test was conducted as "one by one" test where each student fills out the questionnaire individually after trying to apply for the application through the system. The questionnaire had been divided into three parts; demographic, perceived relevancy and perceived usefulness and ease of use. The

sample size was 30 students from University Utara Malaysia. A complete of this questionnaire is included in Appendix A. and we asked them to examine the prototype and provide their feedback. Each of the respondent are asked to use both versions (A and B). The respondent is later required to identify if the result presented by systems are relevant to his/her requirement.

4.4 General information

The General section functions as mechanism to collect student demographics. The System aspects section functions as mechanism to collect data on student opinion regarding to the prototype usability aspects. Figures 4.3, 4.4 and 4.5 summarized Demographic Distribution of the Sample. The Statistical Package for Social Sciences (SPSS) is used to perform descriptive statistics analysis for the collected data. (Coakes & Stead 2007), while the relevancy of the retrieved information is based on precision. (Search & Practices, 1994). The demographic questions include the gender, user degree and field as shown in Figure 4.3, Figure 4.4 and Figure 4.5.

Statistics

N	Valid	30
	Missing	0

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Master	13	43.3	43.3	43.3
	Degree	7	23.3	23.3	66.7
	PH.D	10	33.3	33.3	100.0
	Total	30	100.0	100.0	

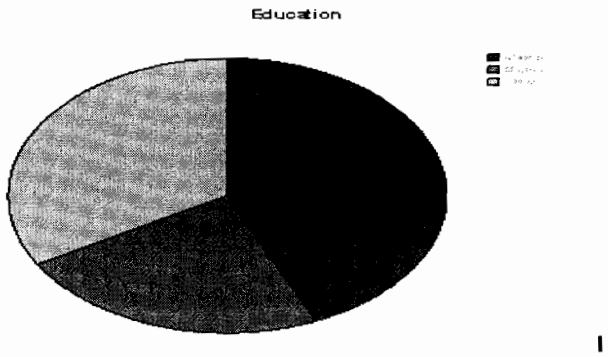


Figure 4.3: Number of Respondent Based on Degree

As shown in Figure 4.3, we have 13 master students and 10 PhD students and 7 undergraduate students whom evaluated the system.

Faculty

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Art & Science	20	66.7	66.7	66.7
	Low	6	20.0	20.0	86.7
	Economic	2	6.7	6.7	93.3
	Banking	2	6.7	6.7	100.0
	Total	30	100.0	100.0	

Statistics

Faculty

N	Valid	30
	Missing	0

Faculty

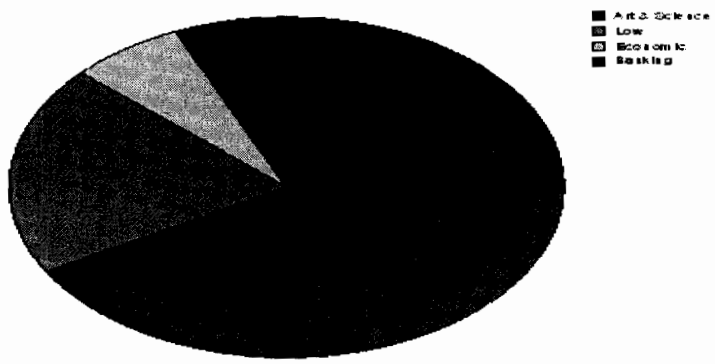


Figure 4.4: Number of Respondent Based on Field

As shown in Figure 4.4, we have twenty students studying in information technology, six students studying in low, two students studying Economic and last two studying Banking .

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	20	66.7	66.7	66.7
	Female	10	33.3	33.3	100.0
	Total	30	100.0	100.0	

Statistics

Gender

N	Valid	30
	Missing	0

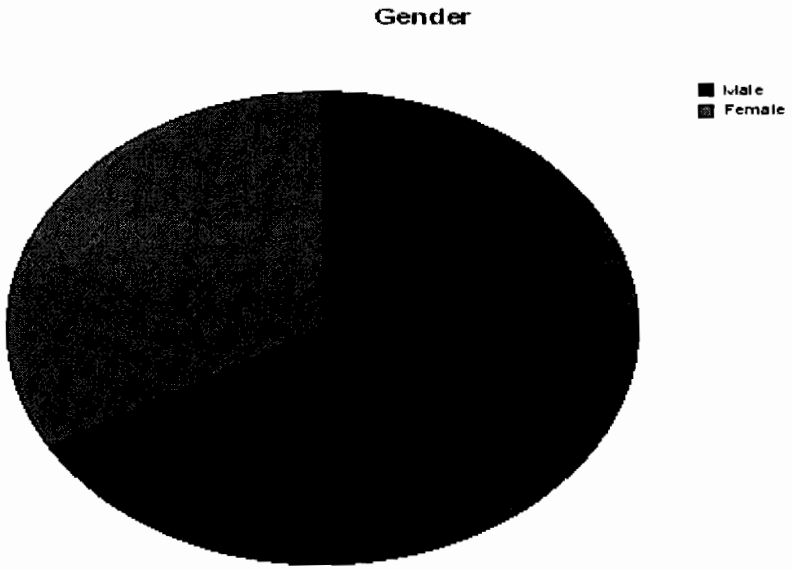


Figure 4.5: Number of Respondent Based on Gender

As shown in Figure 4.3, we have 20 students male and 10 student female

4.4 Analysis of perceived usefulness and ease of use

According to our questions which include main topics to measure perceived usefulness and ease of use, to identify customer satisfaction.

First of all the definition of Perceived ease of use is the degree of ease associated with the use of the system and perceived usefulness is the degree to which an individual believes that using the system will help customers to attain gains in work performance, that's why the questions focus on main four diminution that's reflect usefulness and easy of use as shown in Figure 4.6:

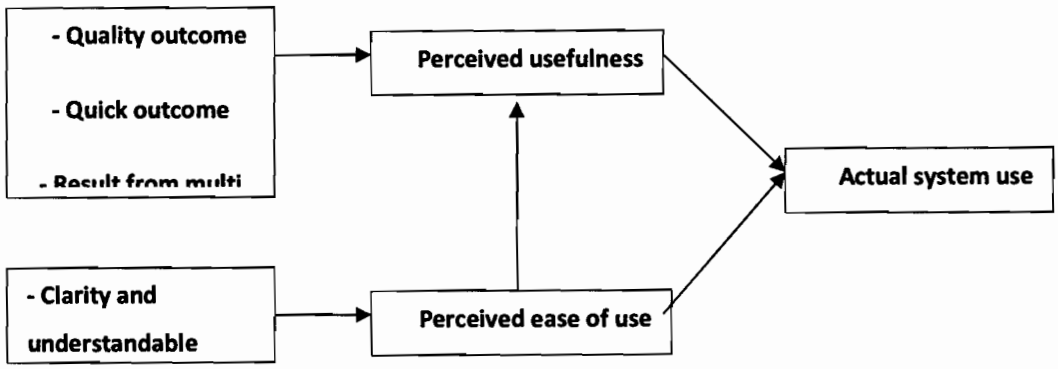


Figure 4.6: Technology Acceptance Model for Second Hand Car

(Source: Davis, et al. 1989)

Special questions designed to measure and identifying all aspect mentioned above which will be analyzed later and discussed in details in this chapter.

In the questionnaire, the evaluation scale had been set as follows:

1. Strongly disagree 2. Disagree 3. Neutral 4=.Agree 5. Strongly agree 6. N/A

As shown in Table 4.1, the total mean for perceived usefulness and perceived ease of use is 4.27775. This value is larger than the mean score (point) which is three and this shows that the respondent tend to accept the proposed approach.

Table 4.1: Descriptive of Perceived Usefulness and Ease of Use

	N	Minimum	Maximum	Mean	Std. Deviation
Perceived Usefulness	30	2.67	5.00	4.2333	.55640
Perceived easy of use	30	3.00	5.00	4.3222	.46719
Total	30			4.27775	

In Table 4.2, we include the number of respondents that correspond to each of the evaluation scale. For example, thirteen students answered strongly agree for the first question which indicates high score in user satisfaction. the second question sixteen students answered strongly agree with a second hand car application that incorporates weighting scheme to represent user preferences which generates a better result compared to one that does not.

As a summary, Table 4.2, reveals that 0.81% (219 answers) answered agrees and (4 answers) answered disagrees with a Concept based term weighting (CBW) application to represent user preferences in information retrieval.

Table 4.2: Questionnaire Likert Scale

Q	Questions	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	N/A
1	Using a car second hand car application that makes use of information from two databases provides a customer with better choices.	0	2	4	11	13	0
2	Second hand car application incorporates weighting scheme to represent user preferences generates a better result compared to one that does not.	0	0	4	9	16	1
3	Using a second hand car application that makes use of information from several databases enable a customer to buy or sale a car more quickly.	0	1	6	10	13	0

4	Using a second hand car application that includes a Concept based term weighting (CBW) with prioritize the user's preferences would improve buy or sale performance.	0	1	4	12	12	1
5	Using the second hand car system for two companies would make it easier to buy or sale a vehicle.	0	0	7	10	13	0
6	I would find a second hand car application with user preferences is useful.	0	0	6	14	10	0
7	Learning to operate the second hand car system would be easy for me.	0	0	3	11	15	1
8	My interaction with the second hand car system would be clear and understandable.	0	0	6	14	10	0
9	I would find a second hand car application easy to use.	0	0	3	14	12	1
Total		0	4	43	105	114	4

4.5 Analysis of perceived relevancy

As presented in chapter three, there are two versions of the system. Version A is build based on hypertext matching while version B applies a Concept based term weighting (CBW). Each of the respondents are asked to use both of versions (A and B). The respondent is later required to identify if the result presented by the systems are relevant to his/her requirement, the precision for twenty seven students is version B results are relevant to them requirement more than version A. the precision for three students is version A results are relevant to them requirement more than version B.we see it in table 4.3 rows 1, 19 and 21, and the precision for student is version A results are relevant to them requirement same version B. we see it in table 4.3 row 4, The respondent is asked to evaluate only ten result presented in the retrieved list. Based on that, we present the precision of the retrieval in Table 4.3. The average perceived relevancy for Version A is 50% and 73% for version B.

Table 4.3: Perceived Relevancy Scale Table

Respondent	Version A	Version B
1	0.7	0.6
2	0.4	0.6
3	0.2	0.7
4	1.00	1.00
5	0.4	0.8

6	0.3	0.7
7	0.2	0.6
8	0.4	0.7
9	0.4	0.7
10	0.3	0.7
11	0.4	0.7
12	0.4	0.6
13	0.4	0.6
14	0.5	0.8
15	0.4	0.6
16	0.4	0.7
17	0.3	0.7
18	0.5	0.7
19	0.7	0.6
20	0.7	0.8
21	0.8	0.7
22	0.8	1.00
23	0.6	0.8
24	0.6	0.7

25	0.5	0.8
26	0.8	1.00
27	0.5	0.6
28	0.6	0.7
29	0.7	1.0
30	0.3	0.7
Precision Average	0.5	0.73

4.6 Summary

Based on the result presented in the earlier subsections, we concluded that the respondents felt version B is better than version A. This shows that by applying a Concept based term weighting (CBW), relevancy of the retrieved results is improved and at the same time fulfilled user preferences.

CHAPTER 5

CONCLUSION

5.1 INTRODUCTION

This chapter focuses on the conclusion and recommendation of this study. The conclusion will explain how this study achieves the goals, according to the problem statements and objectives of this study. Furthermore we have also compared the effectiveness of concept base term weighting approach against one of the existing text retrieved method, namely hypertext-matching in the second hand car trading. Finally, brief recommendations will be given as contributions to future enhancements.

5.2 FINDING

As mentioned in Chapter 1, we have addressed the problem of representing user preferences in retrieving the required information, the main aim of this study is develop and improve information retrieval on a second hand car trading, by adopting a

concept base term weighting to represent prioritize user preferences in identifying the required information such approach has been applied. We have created a second hand car web site, which retrieval Information from two databases. In retrieving the results, we have used two methods namely hypertext method in (system A), and concept base term weighting in (system B), second aim is to evaluate the concept base term weighting approach. For the evaluation first we compare the results retrieved by both approaches (hyper-text matching analysis, concept base term weighting) then, we carried out the usability testing. Based on researcher and users evaluation result presented in chapter 4, we finding that the respondents felt version B is better than version A. This shows that by applying a Concept based term weighting (CBW), relevancy of the retrieved results is improved and at the same time fulfilled user preferences.

5.3 LIMITATIONS OF STUDY

The main problems and limitations of this study (A concept-base term weighting approach to represent user preferences in information retrieval) are:

- We did not get the resources in terms of money, time and distance to visit the second hand car personally.
- The algorithms used in the study are still in evolution process. It was difficult to find the literature supporting the study and results.

5.4 CONTRIBUTION OF STUDY

The contribution of this study can be summarized as follows:

- Conducted A proof of concept study to show the effectiveness of the concept-base term weighting approach.
- Development of a second-hand car trading using C#.NET and SQL server database. That can be sued by users to effecting find or sell their second-hand cars.

5.5 FUTURE WORK

The concept-based term weighting application to represent user preferences in information retrieval make the users query more efficient to help them to solve the problems or the difficulties in the traditional system that uses hypertext matching. Several enhancements can be made on the work which includes:

- A version can be adopted into personal digital assistants (PDAs) and mobile phones.
- Search is based on more than two databases for more choices and profitability.
- The system should work with real lime databases.
- Integrating the developed application with existing online second hand car websites.

5.6 SUMMARY

This chapter summaries the work done during the project time span, findings of the study done, limitations faced during the study, contribution and the recommeded future works.

REFERNCES

Amir, A., Lewenstein, M., & Lewenstein, N., (1997). Pattern Matching in Hypertext, retrieved on September 20, 2009 available at: http://www.cs.biu.ac.il/~amir/Postscripts/hyper.ps.

Black, P. E., (2006) Dijkstra's algorithm, in Dictionary of Algorithms and Data Structures, U.S. National Institute of Standards and Technology, retrieval at September 25, 2009 Available from: http://www.nist.gov/dads/HTML/dijkstraalgo.html.

Blanco, R., & Lioma, C., (2007). Random Walk Term Weighting for Information Retrieval, Amsterdam, Netherlands, SIGIR, ACM, pp. 23-27.

Boyer R.S., & Moore J.S., (2003). Fast string searching algorithm, communications of the ACM, Vol.20, No.10, pp.762-772.

Bogaschewsky, R., & Hoppe, U., (1993). What Hypertext can do for Information Retrieval: Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences, 10.1007/978-0-585-33110-2, Springer Netherlands, pp. 81 - 102.

Brin, S., & Page, L., (2000). The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Science Department, Stanford University, Stanford, CA 94305.

Coakes, S.J, & stead, L. (2007). SPSS version 14.0 for windows. Analysing without English. WILEY.

Colussi L., & Fastest, (1994). pattern matching in strings, Journal of Algorithms, Vol. 16, No. 2, pp 163-189.

Craswell, N. E., (2000), Methods for Distributed Information Retrieval, ©Nicholas Eric Craswell, Typeset in Palatino by TEX and LATEX2e, August 29, 2009; Available: http://research.microsoft.com/users/nickcr/pubs/craswell_thesis00.pdf.

Crochemore, M., & Rytter, W., (1994). The Forward Dawg Matching algorithm, Text Algorithms, Oxford University Press, available at: http://www-igm.univ-mlv.fr/~lecroq/string/fdm.html#SECTION00220

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R., (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, 35, , 982-1003.

Deogun, J. S., Sever, H., & Raghavan,V. V., (1998). Structural Abstractions of Hypertext Documents for Web-based Retrieval, Ninth International Workshop, 10.1109/DEXA.1998.707429, IEEE, pp. 385-390.

Dietinger, T., Gütl, C., Maurer, H., & Pivec, M., (1999). Targeted Information Retrieval, September 16, 2009, Available on: http://www.iicm.edu/iicm_papers/targetedinformationretrieval/ICCE99_target_information.pdf

French, J. C., Knight, J. C., & Powell, A. L., (1997).Applying Hypertext Structures To Software Documentation, Information Processing & Management, Elsevier Science Ltd. Printed in Great Britain 0306-4573/97 $17 +0.00, Pp. 219-231.

Hassan, A. A., (2005). Mixed Heuristic Algorithm for Intelligent String Matching for Information Retrieval, Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05).

Hassan, S., Mihalcea, R., & Banea, C., (2006). Random Walk Term Weighting for Improved Text Classification, In Proceedings of TextGraps: 2nd Workshop on Graph Based Methods for Natural Language Processing, ACL, pp. 53-60.

Hawk Rent A Car, (2008). Kuala Lumpur Car Rental, Kuala Lumpur, Malaysia, retrieval July 25, 2009 available at: http://www.kuala-lumpur.ws/hawk.

Horspool, R.N., (1980) Practical fast searching in strings, Software-Practice and Experience, Vol. 10, No. 6, pp. 501- 506.

Huang, J., Wang, G., & Wang, Z., (2008). Cross-Subject Page Ranking Based on Text Categorization, Zhangjiajie, China, IEEE International Conference on Information and Automation, pp. 363-368.

Islam, Md. R., Sarker, B. D., & Islam, Md. R., (2008). An Effective Term Weighting Method Using Random Walk Model for Information Retrieval, Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, IEEE, pp.1357 - 1362.

Kleinberg, J. K., (1998). Authoritative Sources in a Hyperlinked Environment, Proc. ACM-SIAM Symposium on Discrete Algorithms.

Krovetz R, & Croft W.B., (1992). In proceeding ACM Transactions on lexical ambiguity and Information Systems, Vol 10,2, pp,115-141.

Lopresti, D. P., (1996). Robust Retrieval of Noisy Text, 0-8186-7402-4/96 $5.00 0 1996 IEEE Proceedings of ADL '96, August 29, 2009; Available: http://ieeexplore.ieee.org.eserv.uum.edu.my/iel3/3741/10939/00502518.pdf?tp=&arnumber=502518&isnumber=10939.

Lynch, P., Luan, X., Prettyman, M., Mericle, L., Borkmann, E., & Schlaifer, J., (2004). An Evaluation of New and Old Similarity Ranking Algorithms, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04 $ 20.00, IEEE, pp. 148 – 149.

Mamtani, G., & Green, G., (2006). Reliability Risk Evaluation during the Conceptual Design Phase, Czech Technical University in Prague, Acts Polytechnica Vol. 46 No., pp.8-14.

Mansor, T. A. T., (2007). Tourism. Emerging Malyasia 2007, Oxford Business Group, pp 143-150.

Marchiori, M., (1997). The Quest for Correct Information on the Web: Hyper Search Engines. The Sixth International WWW Conference (WWW). Santa Clara, USA, April 7-11.

Mosteo, A. R., (2006). Dijkstra's algorithm, retrieval at September 25, 2009 available from:Http://coding.derkeiler.com/Archive/Ada/comp.lang.ada/2006-02/msg00138.html.

Nielsen, J. (2000). Designing Web Usability: The Practice of Simplicity. Indianapolis: New Riders Publishing. Retrieved: September 2009, Available from: http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecture notes/ISO9241part11.pdfRanade, A., (1998). A Simple Optimal List Ranking Algorithm, 5th International Conference, IEEE, pp. 60 – 64.

Riloff., E., & Hollaar, L., (1996). Text Databases and Information Retrieval, ACM Computing Surveys, pp.133-136.

Rob, P., & Coronel, C., (2009). Database System: Design, Implementation, and Management, Eight Edition, THOMSON: Course Technology, US.

Roberto Navigli and Paola Velardi(2003). An Analysis of Ontology-based Query Expansion Strategies. In Proceedings of the 14th European Conference on Machine Learning.

Russell, S., (1995). Artificial Intelligence A Modern Approach, Prentice-Hall,Inc, (UK).

Salmeron, J. L., & Smarandach, F., (2007). Redesigning Decision Matrix Method with an indeterminacy-based inference process, September 27, 2009 from: http://arxiv.org/ftp/cs/papers/0703/0703060.pdf

Search, F., & Practices. B., (2006). Demystifying relevance and ranking, UC Berkeley, School of Information Management & Systems,

Spertus, E., (1997). ParaSite: Mining Structural Information on the Web. The Sixth International WWW Conference (WWW). Santa Clara, USA, April 7-11.

Tomassen, S. L., (2006). Research on Ontology-Driven Information Retrieval, August 27, 2008; Available: http://folk.ntnu.no/steint/papers/2006-OTM-ADC-SLT.pdf.

Toms, E. G., (2000). Serendipitous Information Retrieval, September 23, 2009, available on: http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf.

Toms, E. G., (1998). „Information exploration of the third kind: the concept of chance encounters. "A position paper for the CHI 98 Workshop on Innovation and evaluation in Information Exploration Interfaces. (http://www.fxpal.com/chi98ie/submissions/long/toms/index.htm)

Weiss, R., Bienvenido, V., Mark, A. S., Chanathip, M., Peter, S., Andrzej, D., David, K., & Gifford, H., (1996). A Hierarchical Network Search Engine that Exploits

Content-Link Hypertext Clustering. Proceedings of the 7th ACM Conference on Hypertext. New York.

Williams, J, R. (2009). Usability Testing. Retrieved April 1, , from Synergetic Applications web site: http://www.synergeticapplications.com/usability.htm.

Xapian, (2007). Stemming Algorithms, Reviewed on July 8, 2009 avaliabile at: http://xapian.org/docs/stemming.html.

Yan, T. W., Jacobsen, M., Garcia-Molina., H., & Dayal, U., (1996). From user access pattems to dynamic hypertext linking, Computer Networks and ISDN Systems 28 (1996) 1007-1014

Yuwono, B., & Lee, D. L., (1996). Search and Ranking Algorithms for Locating Resources on the World Wide Web, 1063-6382/96, IEEE, pp. 164-171.

Zakos, J., & Verma, B. (2005). Concept-Based term weighting for web information retrieval [Electronic Version]. Retrieved August 9, 2009.

Zeinalipour-Yazti, D., Kalogeraki, V., & Gunopulos, D., (2004). Information Retrieval Techniques for Peer-To-Peer Networks, Computing In Science & Engineering, Pp 20-26.

Appendix A

QUESTIONNAIRE

1. Introduction:

Questionnaire: A Concept based term weighting to represent user preferences in information retrieval. According to Dennis (2002); the second technique is User Testing (questionnaire) to rate the user satisfaction with the web based second hand car application and to evaluate the system in several dimensions.

PLEASE ANSWER THE FOLLOWING QUESTIONS IN ALL PARTS BELLOW:

PART 1: GENERAL INFORMATION:

1. User: Student		
☐ PhD	☐ Master	☐ Undergraduate

2. Faculty :			
☐ FACULTY OF ART AND SCIENCE	☐ FACULTY OF ECONOMICS	☐ FACULTY OF FINANCE AND BANKING	☐ FACULTY OF LOW

3. Gender:	
☐ male	☐ Female

PART 2: SYSTEM ASPECTS:

This part is intended to obtain your views on some aspects of the Second Hand Car.

Please check the appropriate column. The numbers 1 to 5 represent the following:

1 = Disagree, **2** = Strongly Disagree, **3** = Neither, **4** = Agree, **5** = Strongly Agree

1. WAP-BASED USEFULNESS.		**1**	**2**	**3**	**4**	**5**
1	Using a second hand car application that makes use of information from two databases provides a customer with better choices.	☐	☐	☐	☐	☐
2	Second hand car application incorporates Concept based term weighting to represent user preferences generates a better result compared to one that does not.	☐	☐	☐	☐	☐
3	Using a second hand car application that makes use of information from several databases enable a customer to buy or sale a car more quickly.	☐	☐	☐	☐	☐
4	Using a second hand car application that includes a weighting scheme with prioritize the user's preferences would improve buy or sale a vehicle performance	☐	☐	☐	☐	☐
5	Using the second hand car system for two companies would make it easier to buy or sale a vehicle.	☐	☐	☐	☐	☐
6	I would find a second hand car application with user preferences is useful.	☐	☐	☐	☐	☐
1	Using a second hand car application that makes use of information from two databases provides a customer with better choices.	☐	☐	☐	☐	☐
2	Second hand car application incorporates Concept based term weighting to represent user preferences generates a better result compared to one that does not.	☐	☐	☐	☐	☐
3	Using a second hand car application that makes use of information from several databases enable a customer to buy or sale a car more quickly.	☐	☐	☐	☐	☐
4	Using a second hand car application that includes a weighting scheme with prioritize the user's preferences would improve buy or sale a vehicle performance	☐	☐	☐	☐	☐
5	Using the second hand car system for two companies would make it easier to buy or sale a vehicle.	☐	☐	☐	☐	☐
6	I would find a second hand car application with user preferences is useful.	☐	☐	☐	☐	☐

2. PERCEIVED EASE OF USE		1	2	3	4	5
7	Learning to operate the second hand car system would be easy for me.	☐	☐	☐	☐	☐
8	My interaction with the second hand car system would be clear and understandable.	☐	☐	☐	☐	☐
9	I would find a second hand car application easy to use.	☐	☐	☐	☐	☐

Question 10:

Version A:

Kindly check the following results and compare if it relevant to your query or not?

Results	Relevancy
First result	□
Second result	□
Third result	□
Fourth result	□
Fifth result	□
Sixth result	□
Seventh result	□
Eighth result	□
Ninth result	□
Tenth result	□

Version B:

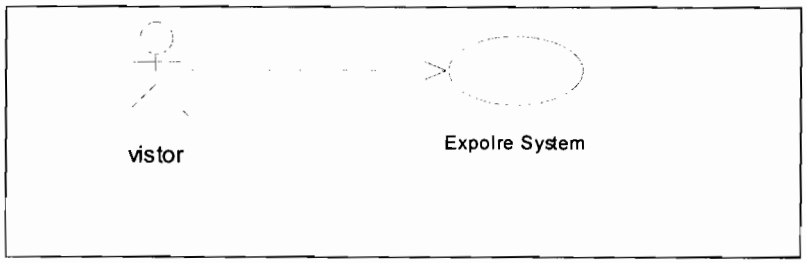
Kindly check the following results and compare if it relevant to your query or not?

Results	Relevancy
First result	□
Second result	□
Third result	□
Fourth result	□
Fifth result	□
Sixth result	□
Seventh result	□
Eighth result	□
Ninth result	□
Tenth result	□

APPENDIX B

USE CASE SPECIFICATION

1. USE CASE: SYSTEM EXPLORE (SHC_01)



1.1 BRIEF DESCRIPTION

This use case is initiated by the visitor. In this use case the visitor can view all the general information about the system (View all company information, View website features, View registration requirements).

1.2 PRE-CONDITIONS

Not Applicable.

1.3 CHARACTERISTIC OF ACTIVATION

Event Driven (on visitor's demand)

1.4. FLOW OF EVENTS

1.4.1Basic Flow (SHC_01_01)

- The visitor can view general information about the system by choice the suitable link from the following links("View company info", "View website features", "View registration requirements ")
- The visitor can print all the previous reports **(A_1: Print)**

1.4.2 Alternative Flow

A-1 Print (SHC _01_08)

The visitor can print all the previous reports by click "Print" button.

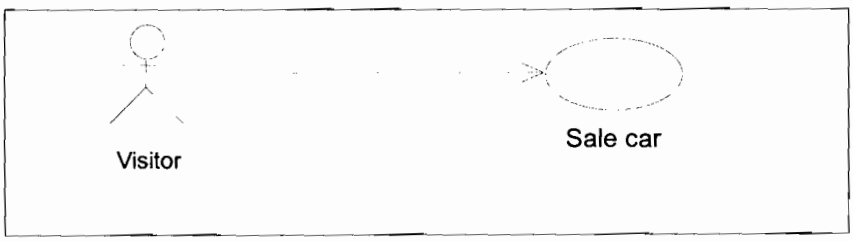
1.5 POST-CONDITIONS

Not Applicable.

1.6 *CONSTRAINT(S)*

Not Applicable.

2. USE CASE: APPLY TO SALE CAR (SHC_02)



2.1 *BRIEF DESCRIPTION*

This use case is initiated by the user. In this use case the visitor can fill the car information application, submit it and the system estimate the price and keep it in the database.

2.2 PRE-CONDITIONS

User must be login already before this action.

2.3 CHARACTERISTIC OF ACTIVATION

Event Driven (on user demand)

2.4 FLOW OF EVENTS

2.4.1 Basic Flow (SHC _02_01)

- This use case begins when the user press the "sale" button.
- The system will display the car info application form panel on the screen.
- The visitors insert the required information in the application fields. **(E-1: Incomplete information).**
- The visitor can press "Submit" button to submit the application.
- The system must estimate the car price.
- The system must keep all the sale application in the database.
- The system will display the car info in the adverts last.
- If the user wants to change the entered information to the default **(A-1: Reset).**
- The visitor can cancel the operation **(A-2: Cancellation).**

2.4.2 Alternative Flow

A-1: Reset. (SHC _02_03)

The visitor can press" RESET "button to back to default information.

A-2: Cancellation. (SHC _02_04)

The visitor can press the "CANCEL" button return back to the main page.

2.4.3 Exceptional Flow

E-1: Incomplete information. (SHC _02_05)

The system will display error message if there is incomplete information.

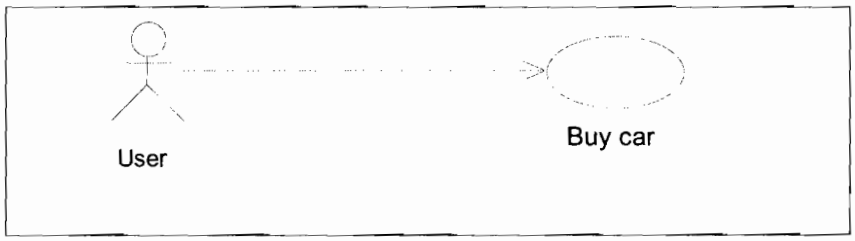
2.5 POST-CONDITIONS

- The system must estimate the car price.
- The system must keep all the sale application in the database.
- The system will display the car info in the adverts last.

2.6 CONSTRAINT(S)

- The user must be registered before.
- the car must be not in the sale last

3. USE CASE: APPLY TO BUY CAR (SHC_03)



3.1 BRIEF DESCRIPTION

This use case is initiated by the user. In this use case the user can fill the buy application, submit it and the system make search in the database and retrieve the information.

3.2 PRE-CONDITIONS

Not Applicable.

3.3 CHARACTERISTIC OF ACTIVATION

Event Driven (on user's demand)

3.4 FLOW OF EVENTS

3.4.1Basic Flow (SHC_03_01)

- This use case begins when the user press the "BUY" button.
- The system will display the buy application form panel.

- The visitors insert the required information in the application fields **(E-1: Incomplete information).**
- The visitor can press "SEARCH" button to retrieve the result.
- The systems display the result.
- If the visitor wants to change the entered information to the default **(A-1: Reset).**
- The visitor can cancel the operation **(A-2: Cancellation)**.

3.4.2Alternative Flow

A-1: Reset. (SHC_03_02)

The visitor can press" RESET "button to back to default information.

A-2: Cancellation. (SHC_03_03)

The visitor can press the "CANCEL" button return back to the main page.

Exceptional Flow

E-1: Incomplete information (SHC_03_04)

The system will display error message if there is incomplete information.

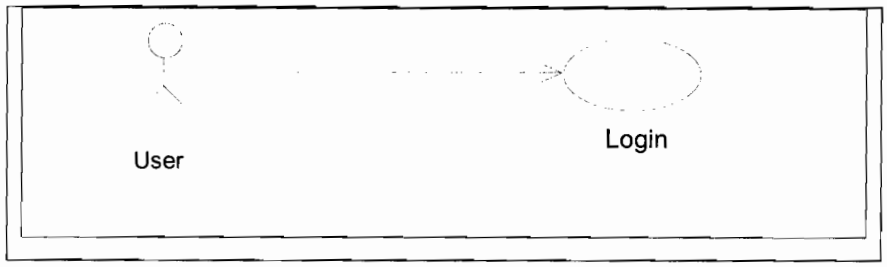
3.5 POST-CONDITIONS

Not applicable.

3.6CONSTRAINT(S)

Not applicable.

14. USE CASE:LOGIN INFORMATION SYSTEM (SHC_04)



4.1 BRIEF DESCRIPTION

This use case is initiated by user. This use case will enable them to use the system to sale his/her car.

4.2PRE-CONDITIONS

The user must have user ID and Password.

4.3CHARACTERISTIC OF ACTIVATION

Event Driven (on user demand)

4.4FLOW OF EVENTS

4.4.1Basic Flow (SHC_04_01)

- The user must enter the login information (user name, password). **(E-1: Invalid Password or user name).**
- The user press "LOGIN" button.
- System must verify the login data.
- The system displays the user (sale) page.

- The user can click "CANCEL" button to return back. **(A-1: Cancellation).**
- The user can click "FORGET" button to retrieve the login information (user name, password). **(A-2: Forget and retrieve)**
-

4.4.2Alternative Flow

- **A-1: Cancellation. (SHC_04_03)**
  The user can press the "CANCEL" button return back to the main page.
- **A-2: Forget and retrieve.(SHC_04_05)**
  The user can click "FORGET" button to retrieve the login information (user name, password).

4.4.3Exceptional Flow

- **E-1: Invalid information (SHC_04_04)**
  The system will display error message if there is incomplete information

4.5POST-CONDITIONS

- The system will allow to the user to login to their main page.

4.6RULE(S)

Not applicable.

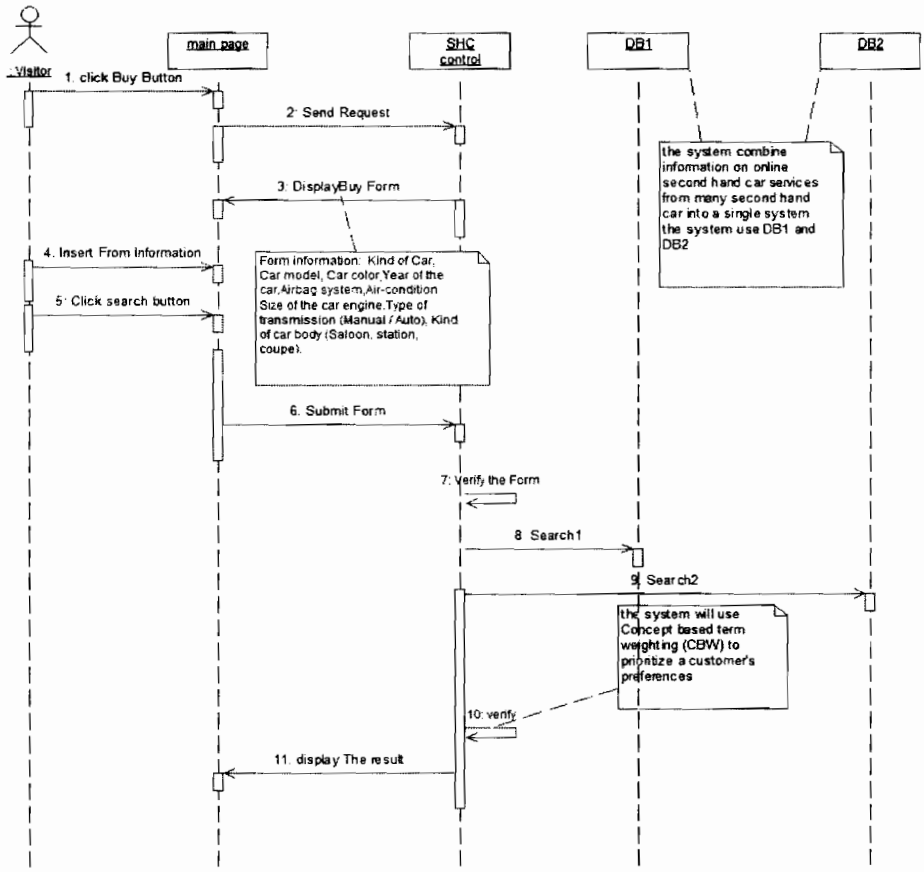
4.7CONSTRAINT(S)

Not applicable.

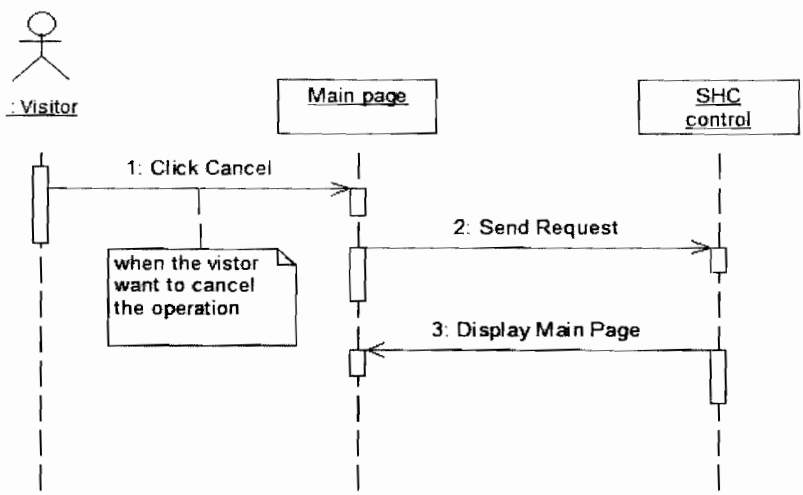
APPENDIX C

SEQUENCE DIAGRAM

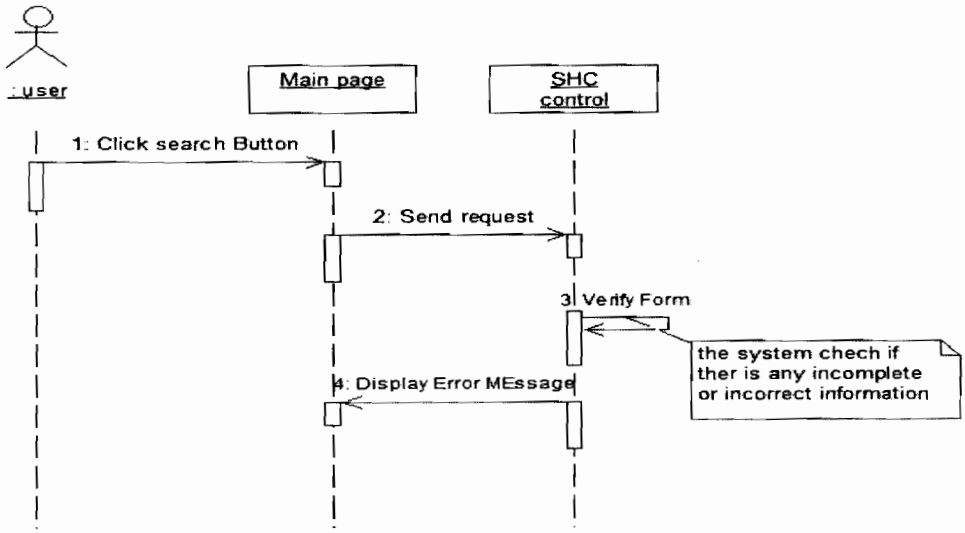
1.1 Apply Buy car (Visitor Basic _ Flow):



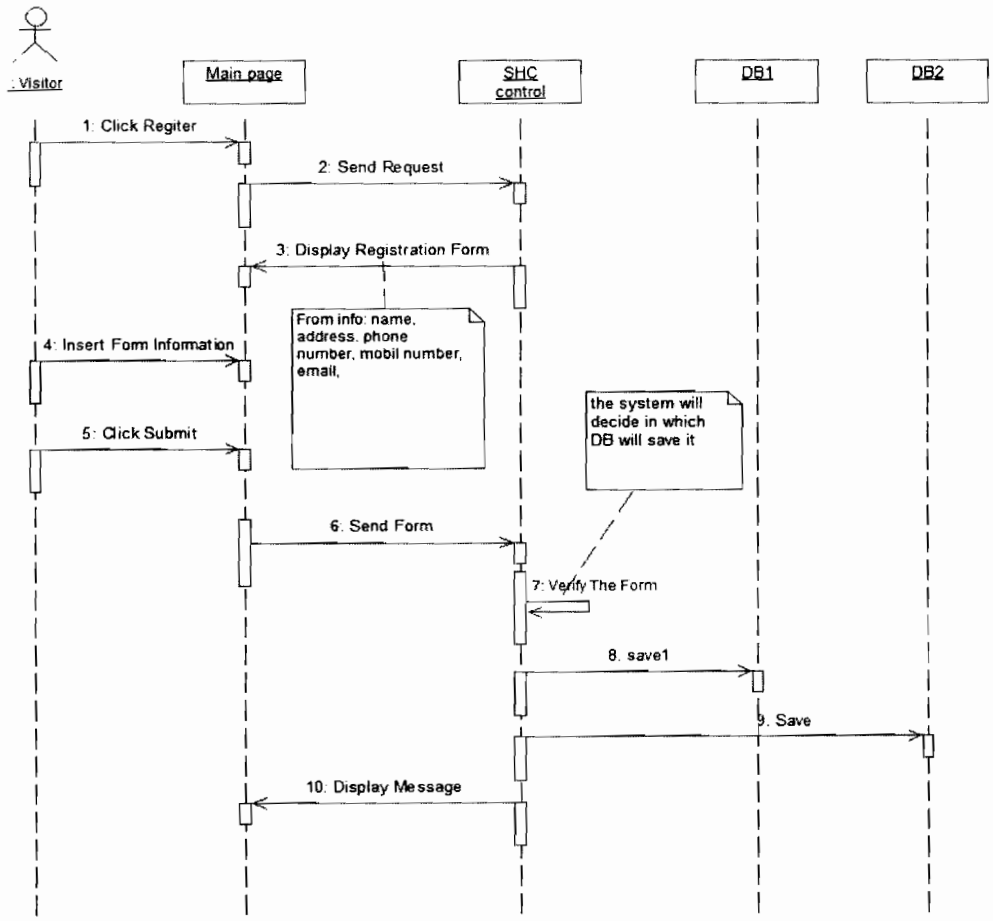
1.2Apply Buy car (Visitor Cancelation _ Flow):



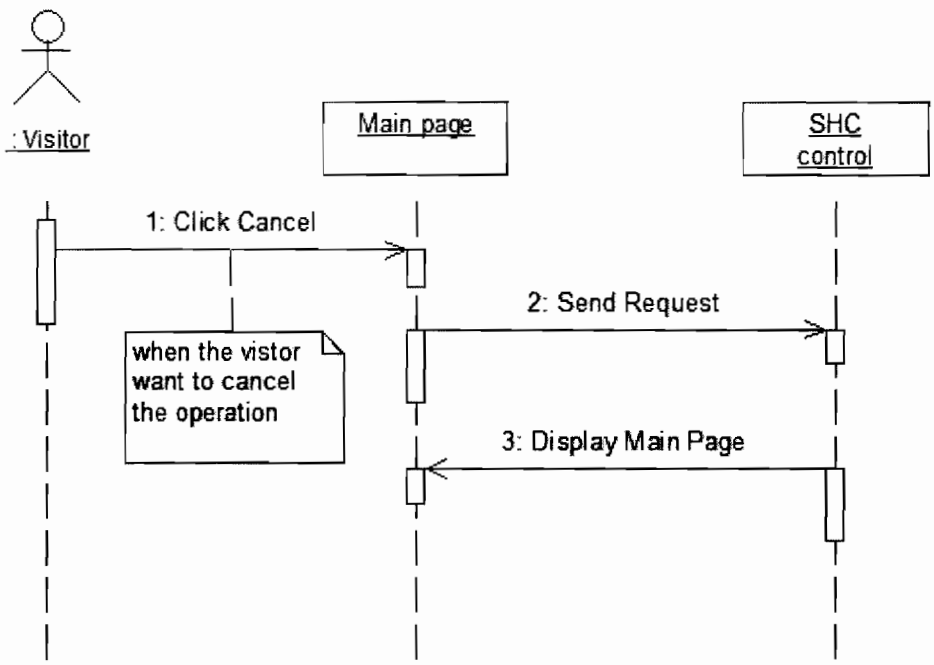
1.3Apply Buy car (Visitor Exception _ Flow):



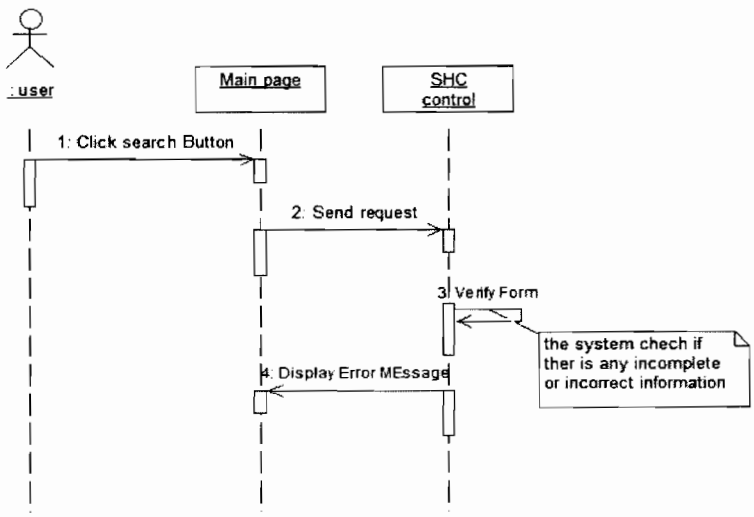
1.4 Apply Registration (Visitor Basic _ Flow):



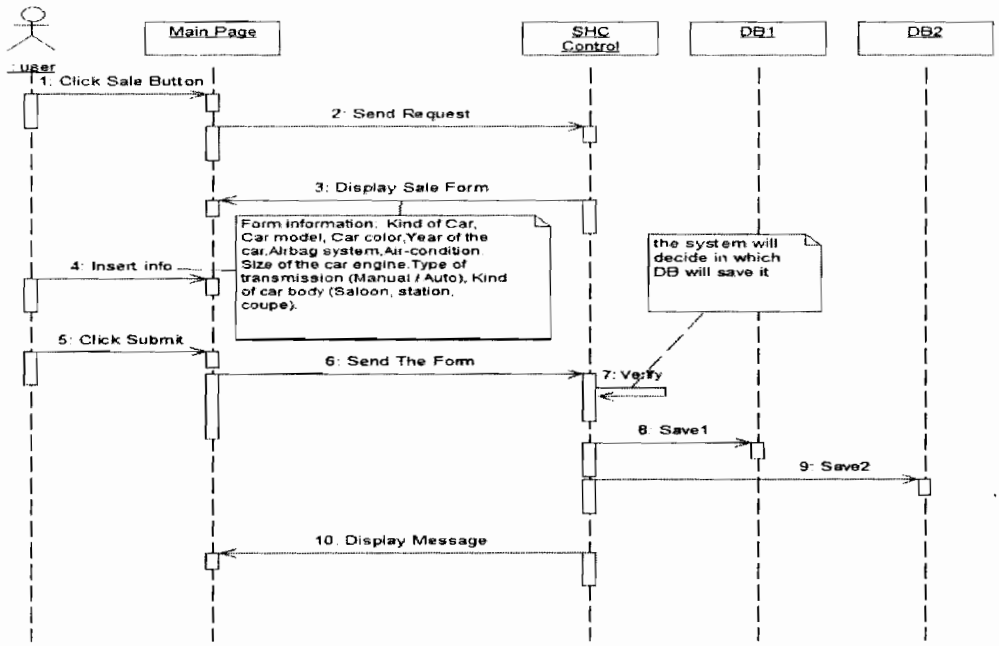
1.5 Apply Registration Policy (Visitor Cancelation _ Flow):



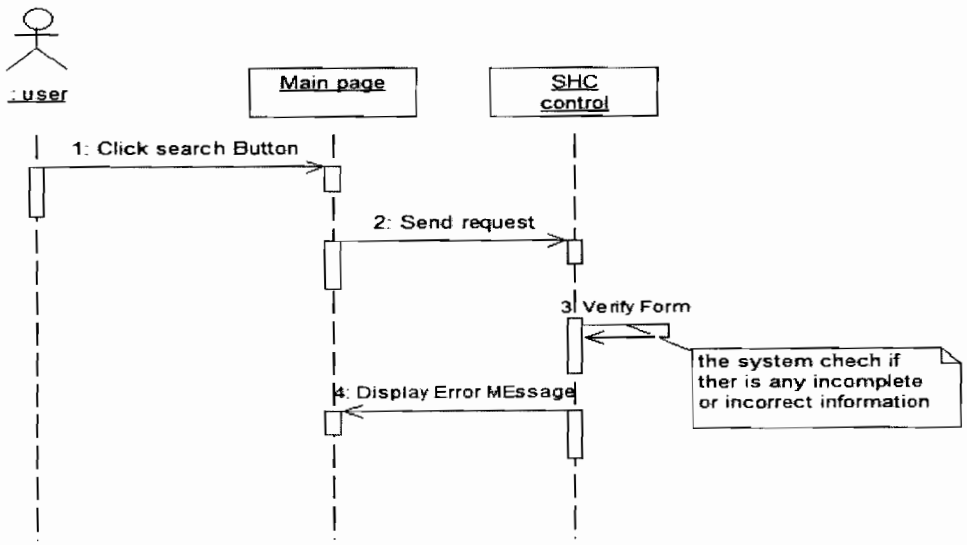
1.6 Apply Registration (Visitor Exception _ Flow)



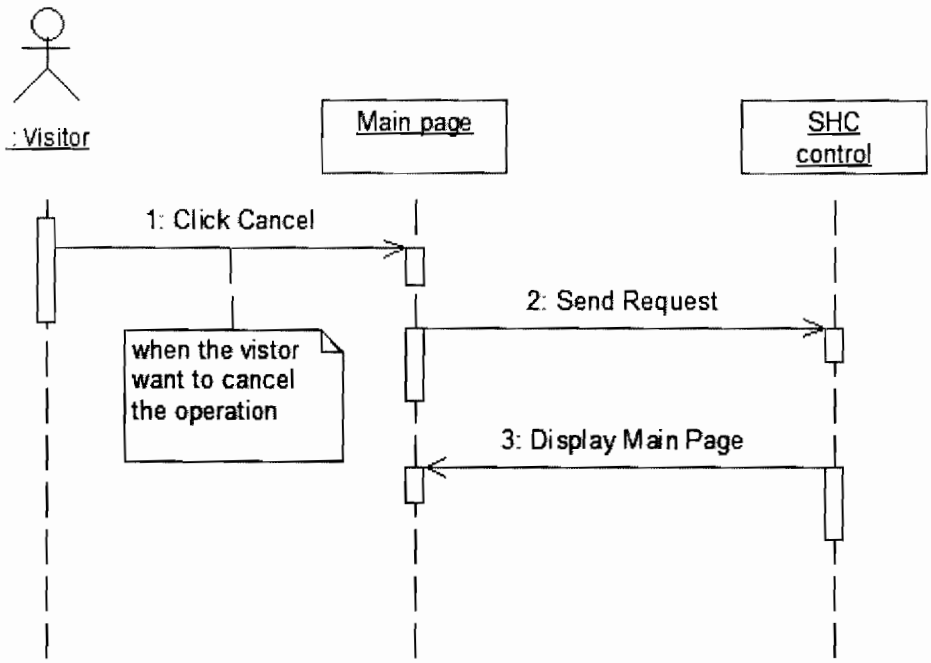
1.7 Apply Sale Car (User Basic _ Flow)



1.8 Apply Sale Car (User Exception _ Flow)



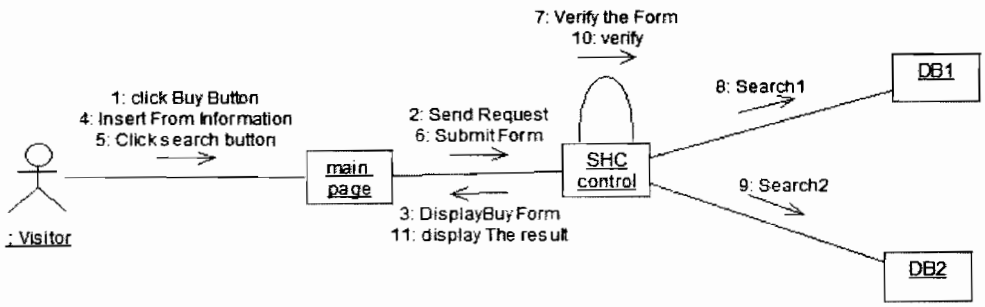
1.9 Apply Sale Car (User Cancelation_ Flow)



Appendix D

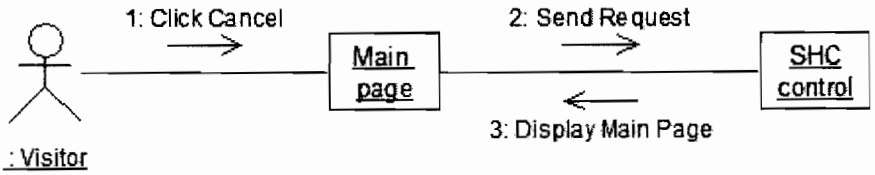
Collaboration Diagram

1. Apply Buy Car (Visitor Basic _ Flow):



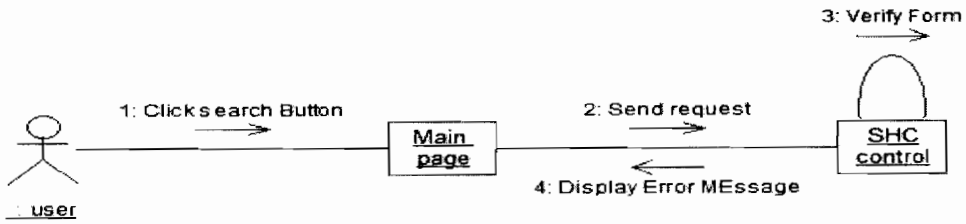
---

2. Apply Buy Car (Visitor Cancelation _ Flow)

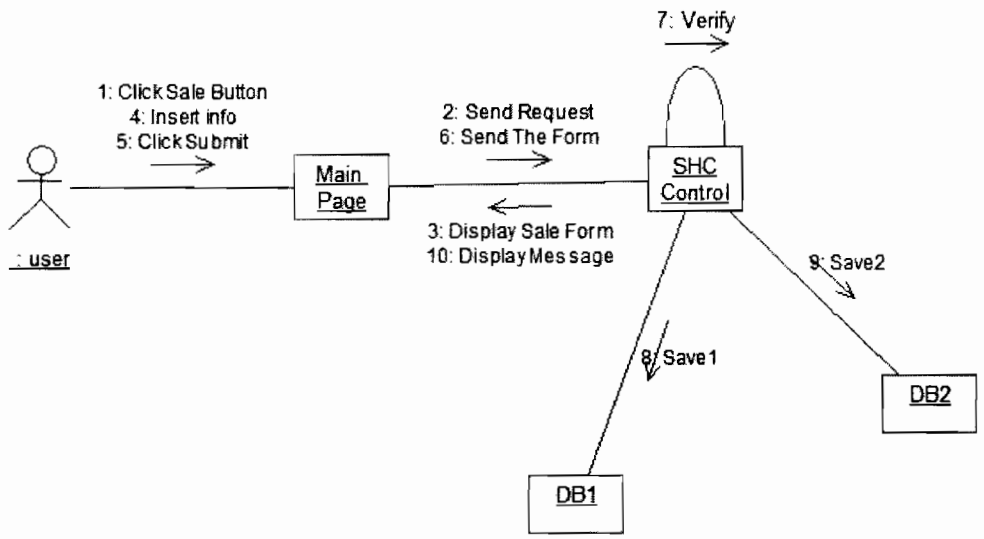


---

3. Apply Buy Car (Visitor Exception _ Flow)

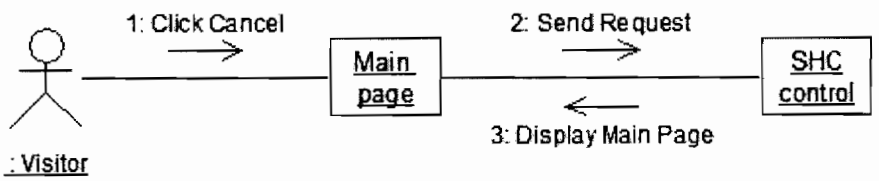


4. Apply Sale Car (User Basic _ Flow):



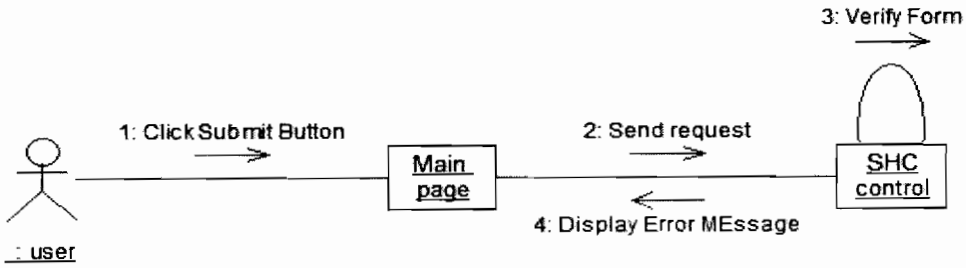
---

5. Sale Car (User Cancelation _ Flow):

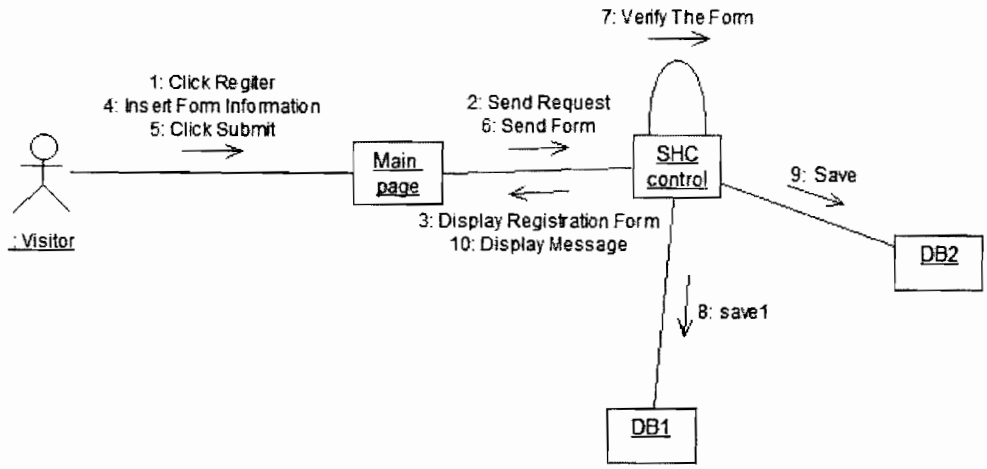


---

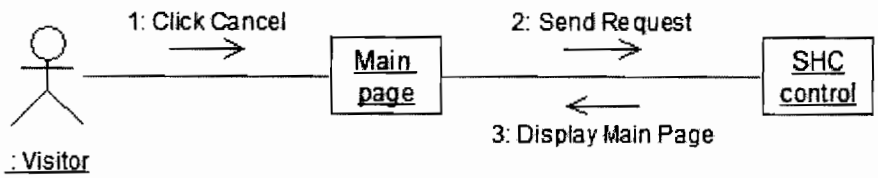
6. Apply Sale Car (User Exception _ Flow):



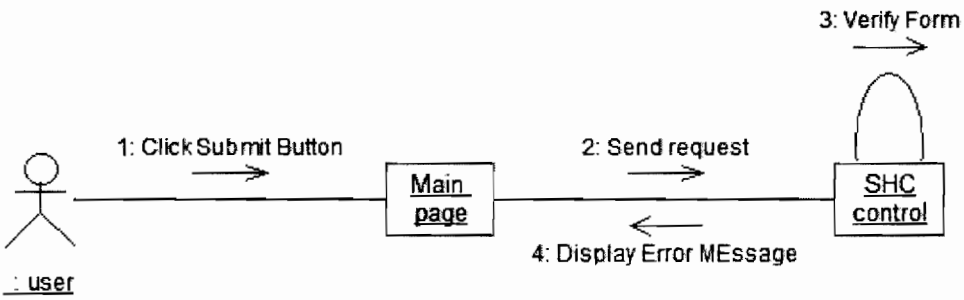
7. Apply Registration (Visitor Basic _ Flow):



8. Apply Registration (Visitor Cancelation_ Flow):



9. Apply Registration (Visitor Exception _ Flow):



Appendix E

Database

Database A

Car information table

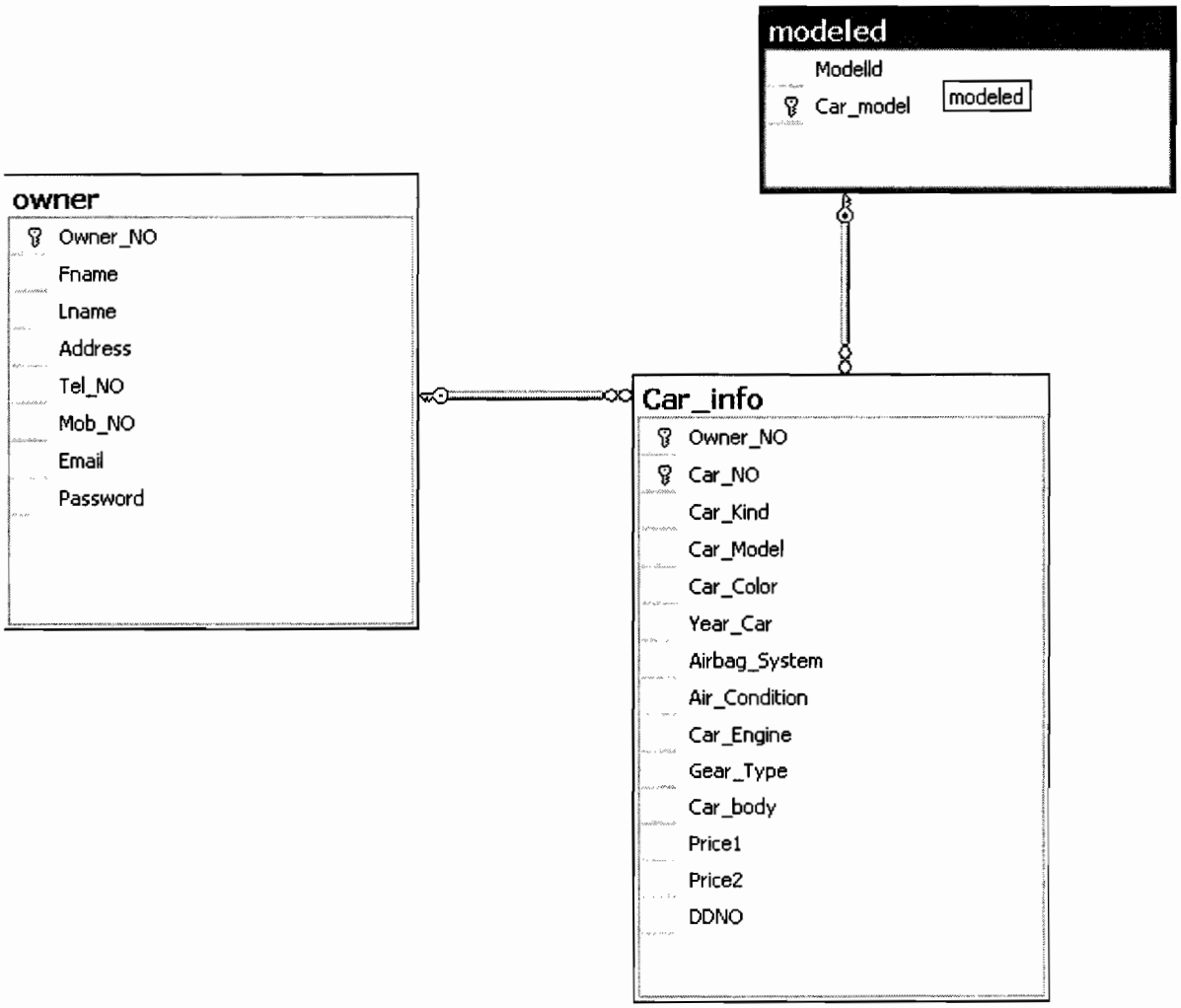
Owner_NO	Car_NO	Car_Kind	Car_Model	Car_Color	Year_Car	Airbag_System	Air_Condition	Car_Engine	Gear_Type	Car_body	Price1
22	222	proton	Gen2	Red	2004	yes	yes	2000	automatic	Station	2000
22	223	proton	Gen2	Red	2004	yes	yes	2000	manual	Saloon	2000
22	225	proton	Gen2	Red	2004	NO	NO	2000	manual	Station	2000
33	3310	proton	Saga	Black	1991	NO	yes	1500	manual	Station	5000
33	3320	proton	Saga	Silver	1991	NO	yes	1500	automatic	Station	6000
33	3330	Toyota	Corila	Red	1985	NO	yes	1300	manual	Station	5000
33	3340	Honda	City	Black	1985	NO	yes	1300	manual	Station	4300
33	3350	proton	Wira	Black	2001	NO	NO	1500	automatic	Station	15800
33	3360	proton	Wira	Blue	2001	NO	yes	1500	automatic	Station	15800
33	3370	proton	Wira	Silver	2001	yes	NO	1500	manual	Station	15800
33	3380	proton	Wira	White	2004	yes	yes	1500	automatic	Station	27000
33	3390	proton	Wira	Black	2004	yes	yes	1500	automatic	Station	27500
55	5510	proton	Iswara	Black	2002	yes	yes	1500	automatic	Station	14300
55	5520	proton	Iswara	Silver	2002	yes	yes	1500	automatic	Station	14500
55	5544	proton	Iswara	Silver	2002	NO	yes	1500	manual	Station	14500
55	5560	proton	Iswara	Of green	2002	yes	yes	1500	manual	Station	13500
55	5570	Toyota	Vios	Black	2004	yes	yes	1500	automatic	Station	53500
55	5580	Proton	Gen2	Cream	2004	yes	yes	1500	automatic	Station	27500
55	5599	proton	Gen2	Black	2004	yes	yes	1600	automatic	Station	27500
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Modeled table

	modelId	Carmodel
▶	4	AUDI
	2	Bmw
	9	Honda
	7	Mazda
	3	mercedes
	5	nissan
	1	proton
	8	Suparu
	6	Toyota
*	NULL	NULL

Owner table

	Owner_NO	Fname	Lname	Address	Tel_NO	Mob_NO	Email	password
▶	[redacted]	bander	Ali	penang	3202045	149052195	bander@yahoo....	ba
	22	bander	ali	penang	123456789	12345678	asdfasdg@yaho...	22
	33	saed	tahat	kedan	233445	23452346	saed@yahoo.com	33
	34	mohand	omar	kedah	3202045	149052195	mohammad	34
	55	omar	omar	penang	14604059	17903920	omar@yahoo.com	55
	66	ahmad	shatnawi	kedah	3098323	149052195	ahmad@yahoo....	66
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL



ERD Database A

Database D

Car Information table

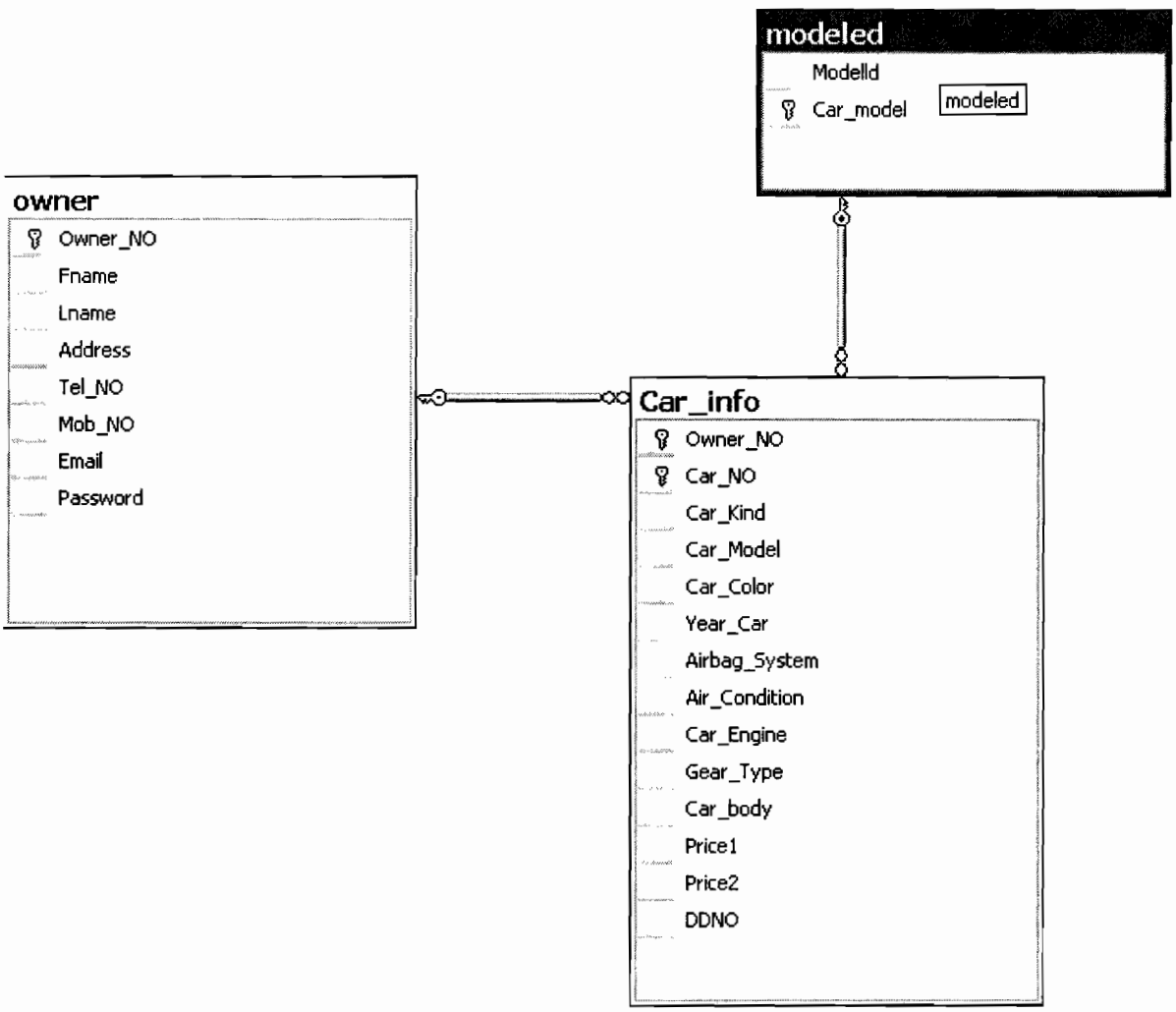
Owner_NO	Car_NO	Car_Kind	Car_Model	Car_Color	Year_Car	Airbag_System	Air_Condition	Car_Engine	Gear_Type	Car_body	Price1
11	112	proton	Iswara	Black	2002	yes	yes	2000	manual	Station	2000
11	113	proton	Iswara	Silver	2003	yes	yes	2000	manual	Station	2000
11	114	AUDI	Wira	Red	2001	yes	yes	2000	manual	Station	2000
11	115	proton	Wira	Black	2001	NO	NO	2000	manual	Station	2000
233	23322	Honda	CRV	Red	2004	yes	yes	2000	automatic	Station	37000
233	23333	Honda	CRV	Black	2000	yes	yes	2000	manual	Station	37000
233	2334	proton	Gen2	Black	2004	yes	yes	1600	manual	Station	26000
233	23344	Honda	Jazz	Black	2003	yes	yes	1400	automatic	Station	61000
233	2335	proton	Gen2	Silver	2004	yes	yes	1600	manual	Station	26000
233	23355	proton	Gen2	Red	2005	yes	yes	1600	automatic	Station	26000
233	2336	proton	Gen2	Silver	2004	yes	yes	1600	automatic	Station	26000
233	23366	Bmw	x5	Red	2005	yes	yes	3000	automatic	Station	159000
233	23377	Bmw	x5	Black	2005	yes	yes	3000	automatic	Station	159000
233	23388	Bmw	x5	Black	2005	yes	yes	3000	manual	Station	159000
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Modeled Table

ModelId	Car_model
4	AUDI
2	Bmw
9	Honda
7	Mazda
3	mercedes
5	nissan
1	proton
8	Suparu
6	Toyota
NULL	NULL

Owner Table

Owner_NO	Fname	Lname	Address	Tel_NO	Mob_NO	Email	Password
11	saed	tahat	kedah	149052196	11323444	asdf@yahoo.com	11
233	Mohammad	shaker	penang	3906040	17845389	mohamd@yahoo...	233
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL



ERD Database B