Recommender System for E-Business Website

by

Syuhada binti Mohamed Ali

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons)
(Business Information System)

SEPTEMBER 2013

Universiti Teknologi PETRONAS
Bandar Sri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Recommender System for E-business Website

by

Syuhada binti Mohamed Ali

A project dissertation submitted to the
Business Information System Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(BUSINESS INFORMATION SYSTEM)

Approved by,

_____
(MR. AHMAD IZUDDIN BIN ZAINAL ABIDIN)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
September 2013

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

_____

(SYUHADA BINTI MOHAMED ALI)

ABSTRACT

The main objective of this project is to provide a recommender system for e-business website that give suggestion of items to the customers based on their preference. The scope of study for this project are the combination of theory of recommender system, e-business website function, analyzing preference of customers based on collaborative filtering algorithms and developing the recommender system for e-business website. A set of questionnaire distributed to a group of people to study more on recommender system from the user's point of view. By using collaborative filtering algorithms, the suggested items of an active user will be predicted based on the ratings of the items by the previous users that have the most similar preferences with the active user. This project using Rapid Application Development (RAD) methodology in developing the prototype as it is the most suitable methodology with the less of 10 months execution time to finish the project. The Gantt chart and key milestones also included to keep track the progress of the project. In results and discussion, the analysis of the survey distributed is discussed and the simulation of recommender system using a web based prototype also discussed in the same chapter. As conclusion, the suggested future work for expansion and continuation in implementing this project also included.

Keywords: Recommender System, E-Business Website, Collaborative Filtering.

ACKNOWLEDGEMENTS

"Praise to Allah, the most Gracious and the most Merciful''

Alhamdulillah praises to Allah for the strengths and guidance that He gave to me to overcome all the challenges and problems in completing my final year project. My deepest gratitude also to Him because with His blessing I can complete my journey for these 28 weeks.

I would like to extend a special thanks to my supervisor, Mr. Ahmad Izuddin bin Zainal Abidin for providing me guidance during these two semesters. His patience in guiding me and his determination to make me learn as much as I can is highly appreciated. He always supports me along the way since I attached under his supervision and he also always lending his ears to hear my problems and give some advises to me.

I would also not forget Azfar Tommy, post graduate student at Computer Information Science department that helps me a lot in completing my final year project. Hope we can continue our friendship and cooperation in the future. Nevertheless, sincere thanks to all my friends who always helped me and believed that I can complete my final year project especially Nur Syafiqah, Sharifah Sarah, Aniz Syuhaiza, Puteri Intan Aishah and Noor Hasomy. I will never forget the time that we were together, as well as sharing our experience together, thank you again.

Last but not least, my sincere appreciation goes to my beloved parents, Mr. Mohamed Ali bin Abdul Rahim and Mrs. Saadiah binti Md Din. Not to forget my sister and brothers, Sharizan, Azhar and Abdul Razak for their endless love, prayers and encouragement. To Mohammad Haris bin Hazizan, thank you very much. To those who indirectly involved in my final year project, your kindness means a lot to me. Thank you to all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND NOMENCLATURES

CF Collaborative Filtering
RAD Rapid Application Development

CHAPTER 1
INTRODUCTION

This chapter will describe the overview of the project that covers the topics below.

- Background of Study
- Problem Statement
- Objectives of Study
- Scope of Study
- Relevancy and Feasibility of Study

1.1 Background of Study

E-business is a way of doing business on the part of internet called the World Wide Web (the web). It started in the mid-1990s and it grew rapidly until 2000, when a major economic downtown occurred. It rose back in the beginning of 2003, and it continued to grow during the recession of 2008-2009 (Schneider, 2011). The increasing number of people who have access to the internet helps e-business to successfully established and become a phenomenon. Some businesses are click and mortar, which means it operates in a physical store and also has a website (e-business) to operate their businesses. Some of the businesses only existed virtually which means totally using e-business doing the business transactions.

When it comes to e-business, people will browse the website to buy their desired item. However, some of the people do not know what items to choose or need some suggestion which items to buy. Hence, recommender system is use to help their potential customers to find the most suitable items according to their preference. It is a tool that will give suggestion to the potential customers which items is mostly choose by the previous customers that have the most similar preference like the potential customers.

1.2 Problem Statements

i. Some of the customers have problems in deciding which items to buy because the e-business website provides too many information about the items. The customers need to read a long description about the items to get to know more about the items.
ii. E-business website usually has a static or dynamic product catalog to display what products are available or sell in their website. However, it only shows the items according to the category that set by the business. It did not help the customers to find complimentary products that usually used with the main product that they want to buy.
iii. Some of the businesses that go to online did not exist in a long time because they cannot maintain their customers. It is because they cannot attract their customers to buy their new products and they cannot maintain their sales.

1.3 Objectives

i. To provide a recommender system for e-commerce website that give suggestion of items to the customers based on their preference.
ii. To generate a list of suggested complimentary items to be use by the customers based on the main items that they want to buy.
iii. To help e-business website to maintain their customers by giving a value of appreciation by knowing the preferences each of their customers.

1.4 Scope of Study

i. Theory of recommender system.
ii. E-business website.
iii. Analyzing preference of customers based on collaborative filtering (CF) algorithms.
iv. Finding the most feasible way in developing recommender system for e-business website.

1.5 Relevancy and Feasibility of Study

The purpose of this study is to explore the theory of recommender system in order to develop an effective recommender system for e-business website that will consequently address the problem statements. This research will later be proved on a website that demonstrates the process of suggesting most suitable items for the customers. Other than that, these studies also will persistently the e-business website to generate revenue by using the recommender system. Furthermore, e-business website is growing rapidly in Malaysia because it is an easier way for the businesses to go global and have more customers.

The time frame of the project development divided into two semesters of study; whereby the first semester (during FYP I: May 2013 – August 2013), the project focused on extensive research, planning and analysis phases, while the second semester (during FYP II: September 2013 – January 2014), the project mainly focused on designing, modeling, prototyping and developing the recommender system for e-business website. Other than that, the time frame given is feasible and practical with the scopes of study for this project.

CHAPTER 2
LITERATURE REVIEW

This chapter discusses on the general idea of the project. It contained the background information on recommender system and also identified what others have said and discovered about the recommender system. It also provides certain theories and facts that relevant to the objectives and findings of this project.

2.1 Data and Knowledge Sources

According to Ricci, Rokach and Shapira (2011), recommender system is an information processing systems that actively and gathers variety types of data in order to provide the recommendations. Data are combinations of the items to suggest and the users who will receive these recommendations. Data used by the recommender system can be classified into three objects:

i. *Items*

Items refer to the objects that recommended. It can be characterized based on their complexity and their value or utility. If the item is useful to the user, it can be label as positive, and the item is negative when the item is not appropriate for the user and the user made a wrong decision when selecting it. When a user browse the website to find their desired items, a user will always incur a cost that includes cognitive cost of searching for the item and the real monetary cost to pay for the item. Example of items with low complexity and value are news, Web pages, books, CDs, movies. Example of items with larger complexity and value are digital cameras, mobile phones, PCs, etc.

ii. *Users*

Users of a recommender system may have different objectives and characteristics that different from each other. In order to personalize the recommendations for every single user, recommender system manipulates the

information about the users. This information can be structured and arranged in many ways and the selection of what information to model is based on the recommendation techniques used. As example, in a demographic recommender system, socio-demographic attributes such as age, gender, profession and education are used.

iii. *Transactions*

Transactions can be defined as a transaction that recorded interaction between a user and the recommender system. Transactions usually are log-like data that store important information generated during the human-computer interaction and useful for the recommendation generation algorithm that a system used. It also may contain references to the item selected by user and a description of the content for that particular recommendation. Other that that, transaction also included the explicit feedback from the user, such as the rating of the selected item. According to Schafer, Frankowski, Herlocker and Sen (2007), there are various types of ratings which are:

- Numeral ratings such as the 1-5 stars provided in the book recommender associated with Amazon.com.
- Ordinal ratings, such as "strongly agree, agree, neutral, disagree, strongly disagree" where the user is required to select the term that best indicates her or his opinion regarding an item and its usually held via questionnaire.
- Binary ratings that model choices in which the user is simply need to decide if a certain item is good or bad.
- Unary ratings can indicate that a user has observed or purchased the item, or rated the item positively. Generally in such cases the absence of a rating indicates that we have no information relating the user to the item and perhaps she or he purchased that item somewhere else.

2.2 Recommendation Techniques

A common strategy for modeling user in the recommender system is for users to assign rating to the items. When a target user or the potential customers wants recommendations, recommender system will calculate predictions and estimates of how the target user would rate the items. For the next target users, the system will recommend the items that have high ratings (Cosley, Lam, Albert, Konstan & Riedl, 2003). As mentioned by Peng, Bo, Fan and Ruimin (2004), two techniques that commonly used in the recommender system are content-based filtering which using the content of the recommended items, and also collaborative filtering which using ratings of similar users on the recommended items. However, collaborative filtering has proved to be one of the most successful techniques in recommender system, with the key idea of the users will prefer items with people that have same preferences.

Collaborative filtering recommender system model consist of the degree of utility of the user u for the item *i* as a (real valued) function *R(u,i)* by considering the ratings of users for the items. Collaborative filtering technique is to predict the value of *R* over pairs of users and items, i.e., to compute $\hat{R}(u, i)$, where R as the estimation, computed by the recommender system, of the true function *R*. By using active user u on a set of systems, i.e., $\hat{R}(u, i_1), \ldots, \hat{R}(u, i_N)$ the recommender system will suggested the items. They model this degree of utility of the user *u* for the item *i* as a (real valued) function $R(u, i)$, as is normally done in collaborative filtering by considering the ratings of users for items.

Then the fundamental task of a collaborative filtering recommender system, is to predict the value of *R* over pairs of users and items, i.e., to compute $\hat{R}(u, i)$, where we denote with $\hat{R}$ the estimation, items $i_{j1}, \ldots, i_{jK}$ ($K \leq N$) with the largest predicted utility. Usually *K* is a small number, i.e., much smaller than the cardinality of the item data set or the items on which a user utility prediction can be computed (Ricci, Rokach & Shapira, 2011).

2.3 Recommender System Function and Benefits

Recommender system is highly aimed for the potential customers that lack of personal experience or competence to evaluate and choose the potentially overwhelming number of alternative items in the e-business websites. The role of the recommender system provider which is the (e-business website) must be distinguished with the user (customers) of the recommender system. For example, a travel recommender system is normally introduced by a travel intermediary or a destination management organization to increase their revenue by increase the numbers of tourists go to the vacation destinations (Ricci, Rokach & Shapira, 2011). As also mentioned by Ricci, Rokach and Shapira (2011), there are many benefits of recommender systems which are:

i. *Increase the number of items sold.*

   This the most essential function of recommender system which is to be able to sell an additional set of items compared to items that usually sold without any kind of recommendation. Recommended items are the items that mostly match with the user's needs and wants. From the point of view of the recommender system provider, recommender system is introduce to increase the conversion rate, for example, the number of users that convert to customers and consume an item, not just visitors that only browse through the information but did not buy the items.

ii. *Sell more diverse items.*

   Recommender system allows the users to choose items that not easily to find without a precise recommendation.

iii. *Increase the user satisfaction.*

   Experience the user of the e-business website can also improve with a well designed recommender system. The combination of effective, interesting, relevant and, with a properly designed human-computer interaction will increase the user's attention to the e-commerce website.

iv. *Increase user fidelity.*

A user will feel appreciated when the website that they visited recognizes them as old customer and treats him as a valuable visitor. Leveraging the information acquired from the user in previous interactions such as his ratings of items is one of the features of the recommender system. As also mentioned by Cosly, Lam, Albert, Konstan and Riedl (2003), a good recommender system can inspire trust towards the business and help users find the products that they truly want.

The algorithm used in the recommender system can be changed and exploit in order to provide a better recommender system. Other than that, by fully used the characteristics of the items and ratings from the previous users in generating the suggestion items for the potential customers will produced a successful recommender system for an e-business website.

CHAPTER 3
METHODOLOGY

It is important to choose the right methodology in developing the products of the research study; which are Web pages with the recommender system in it. Different methodologies could cater different needs of a project in a period of time. Thus, this chapter will discuss on:

- Product Development Phases
- Project Activities
- Tasks Schedule and Key Milestones
- Tools for Development
- Mathematical Formulas/Algorithms
- System Architecture

3.1 Product Development Phases

In order to completing this project, the main methodology used is the Rapid Application Development (RAD) method. This methodology suitable for faster development and it give a better quality results to compare with the traditional methodology. Rapid Application Development (RAD) method takes an approach whereby minimal planning is required and it focused more towards rapid prototyping. This methodology also suitable with the time constraint to complete this project which is less than 10 months. There might be possibilities of functionality and performance of the system will be compromised and modified along the development process. Thus, this flexibility enables this project to complete within the time frame given.

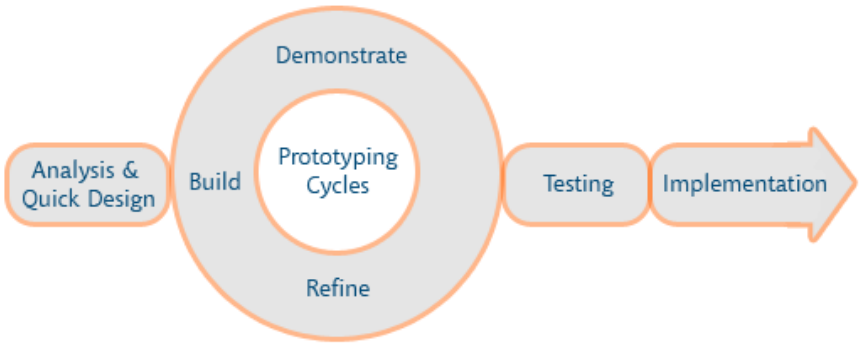


FIGURE 1: Rapid Application Development (RAD) Methodology

3.2 Project Activities

This methodology provided four main phases in order to ensure the project progress meets the key milestones that have been set earlier. The tasks involved are shown in the table below.

TABLE 1: Project Activities

Phase	Project Activities
Analysis and Quick Design	• Literature review and research on the subject matters (Recommender system theory, algorithms, etc). • Prepare set of questionnaire for data gathering and carry out survey. • Sketch brief test plan. • Analyzing the data that gathered from the survey by using basic statistics. • Prepare the propose recommender system architectural design. • Prepare Gantt chart and key milestones for task accomplishment and progress, and also to mark important points in the development process.

Prototyping Cycles	*i. Build* • Prototyping the user interface and Web pages contents according to the test plan sketched. • Provide dataset. • Write codes and develop Web pages contents with the recommender system. *ii. Demonstrate* • Run simple test to show the workability of the prototype. • Ensure all the components in the prototype interrelated and working. • Publish Web pages on live server. *iii. Refine* • Reconstruct the prototype for improvement purpose, after reviewing the system faults of the previous version. • Revert back to the first and second prototyping cycles (build and demonstrate) if needed, until the system meets the requirements and objectives.
Testing	• Evaluate the prototype functionality and usability on live server. • Check the specifications of the prototype whether they are aligned with the requirements and meet the objectives of the study.
Implementation	• The prototype is ready to be used when it has passed all testing phase. • Provide further recommendations for the system.

3.3 Gantt Chart and Key Milestones

All main tasks in completing is project have been tabulated in a Gantt chart for a better view. It helps to visualize the current progress of the project. Below is the task schedule according to the time frame given by the coordinator of Final Year Project II of Computer Information Sciences (CIS) department.

TABLE 2: Gantt Chart and Key Milestones

First Term: Final Year Project I															
No	**Tasks**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Analysis and Quick Design															
1	Literature review and research on the subject matters.														
2	Prepare survey questionnaires for data gathering and carry out survey.														
3	Sketch brief test plan.														
4	Analyzing the data that gathered from the survey.														
5	Develop the architectural design.														
6	Develop the recommender system flow chart.														
7	Prepare Gantt chart for task accomplishment and progress.														
8	Develop key milestones that mark significant points in the development process.														

Final Term: Final Year Project II															
No	**Tasks**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Prototype Development															
i-Build															
9	Prototyping the user interface and Web pages contents according to the test plan sketched.														
10	Provide dataset.														
11	Write codes and develop Web pages contents with the recommender system.														
ii-Demonstrate															
12	Run simple test to show the workability of the prototype.														
13	Ensure all the components in the prototype interrelated and working.														
14	Publish Web pages on live server.														
iii-Refine															
15	Reconstruct the prototype for improvement purpose, after reviewing the system faults of the previous version.														

16	Revert back to the first and second prototyping cycles) if needed, until the system meets the requirements and objectives.														
Testing															
17	Evaluate the prototype functionality and usability on live server.														
18	Check the specifications of the prototype whether they are aligned with the requirements and meet the objectives of the study.														
Implementation															
19	The prototype is ready to be used when it has passed all testing phase.														
20	Provide further recommendations for the system.														

3.4 Tools for Development

Tools that have been used for this project are:

i. WampServer

WampServer is a Windows web development environment. It allows the creation of web applications with Apache2, PHP and MySQL database. Alongside, PhpMyAdmin allows the managing of databases in easy way.

ii. Apache HTTP Server

Apache HTTP server is an open source Web server platform. This tool is utilized to deliverer Web pages through the Internet. When the Apache program receives request for a file, it will looks for the file on its disks, and when it found, it will sends that file to the requester in HyperText Transfer Protocol (HTTP).

iii. Hypertext Preprocessor (PHP)

PHP is a server-side scripting language designed for web development. It will generate the resulting Web page. Other that that, PHP also can be embedded directly into an HTML source document.

iv. MySQL

MySQL is a database system used on the web that runs on the server. It is ideal for both and small large applications and very fast, reliable and easy to use. It also supports standard SQL and also can compiles on a number of platforms.

3.5 Algorithms

As discussed in chapter 2, collaborative filtering has proved to be one of the most successful techniques in recommender system, with the key idea of the users will prefer items with people that have same preferences. User-based CF is a memory-based algorithm which tries to mimics word-of-mouth by analyzing rating data from many individuals. The assumption is that users with similar preferences will rate items similarly. According to Hahsler (2011), the missing ratings for a user can be predicted using:

i. Cosine similarity algorithm - Finding a neighborhood of similar users.
ii. Predicted ratings - The aggregate ratings of the users in the neighborhood to form a prediction.

3.5.1 Cosine Similarity Algorithm

The neighborhood can be define in term of similarity between users, either by taking a given number of most similar users which is called *k* nearest neighbors, or all users within a given similarity threshold. The similarity measures are defined between two users, active user and existed users. One of the ways to calculate the similarity is by using the Cosine similarity algorithm. (Hahsler, 2011)

$$w(a,i) = \sum_j \frac{v_{a,j}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2}} \frac{v_{i,j}}{\sqrt{\sum_{k \in I_i} v_{i,k}^2}}$$

$$\mathrm{sim}_{\mathrm{Cosine}}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

FIGURE 2: Cosine Similarity Algorithm

Basically, *va,j* refer to vote of active user for item j and *vi,j* refer to vote of user i (existed user) for item j. All this ratings will put in the rows and columns in the utility matrix. In simple way, the similarity measures are defined between two users, active user (x) and existed user (y). Where, x = row x and y = row y represent the row vectors in the utility matrix with the two users' profile vectors. For calculating similarity using rating data only the items are used which were rated by both users.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	?	4.0	4.0	2.0	1.0	2.0	?	?
u_2	3.0	?	?	?	5.0	1.0	?	?
u_3	3.0	?	?	3.0	2.0	2.0	?	3.0
u_4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
u_5	1.0	1.0	?	?	?	?	?	1.0
u_6	?	1.0	?	?	1.0	1.0	?	1.0
u_a	?	?	4.0	3.0	?	1.0	?	5.0
r_a	3.5	4.0			1.3		2.0	

FIGURE 3: Utility Matrix

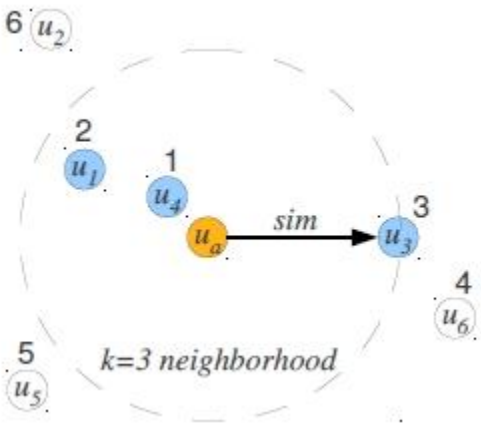


FIGURE 4: User Neighborhood Formation

As we can see in the figure 3, it is the rating matrix R with 6 users and 8 items ratings in the range 1 to 5 (stars). The recommendation is for the active user *ua* which located at the bottom of the matrix. To find the *k*-neighborhood or the *k* nearest neighbors, the similarity between the active user and all other users calculated based on their ratings in the database using Cosine similarity algorithm and then select *k* users with the highest similarity. The average value of Cosine similarity of the active user and with each of the user needs to be computed to find the *k*-neighborhood. In the figure 4, a 2 dimensional representation of the similarities, which means user with higher similarity are displayed closer with the active user in the center. The $k = 3$ nearest neighbors are *u4, u1* and *u3* are selected and marked in the database to the left.

3.5.2 Predicted ratings

Once the users in the neighborhood are found, their ratings are aggregated to form predicted rating for the active user. To generate an aggregated predicted rating, the average ratings in the neighborhood for each item not rated by the active user need to be computed. To create top-*N* recommendation list, the items are ordered by predicted rating. In figure 3, the predicted ratings computed is 3.5, 4.0, 1.3 and 2.0 for item 1, item 2, item 5 and item 7 respectively. Based on this value, the top-*N* recommendation list is item 2, item 1, item 7 and item 5.

3.6 System Architecture

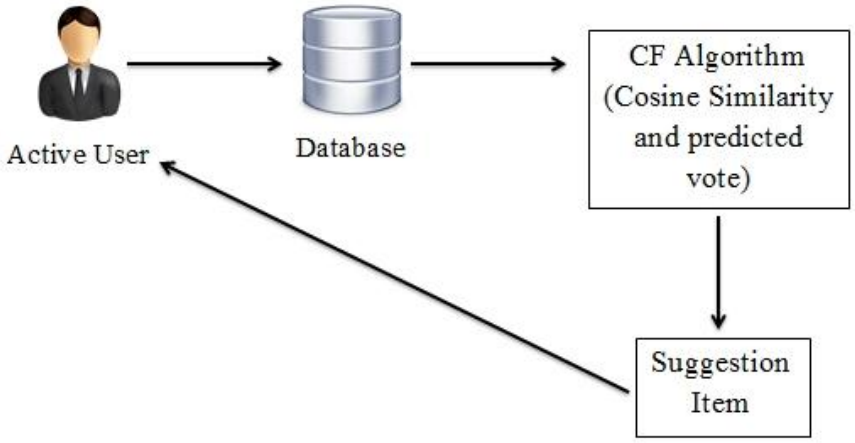


FIGURE 5: System Architecture

All users need to give ratings for all the items that they buy, at least for two items. It is because in order two compute the Cosine similarity algorithm, it will use two vectors which are for active user and existed user, with at least two similar items that both of the users have rated. All of the user ratings will be stored in the database. When an active user need recommendation, the ratings for the items by the active user and existed users will be extracted from the database. Then, the data is use to calculate the Cosine similarity and the predicted ratings. The system will give the list of recommended items to the active user based on the calculation. The item that has the highest predicted ratings is the most recommended item for the active user.

CHAPTER 4
RESULTS AND DISCUSSION

This chapter is a core component of the whole progress report whereby results and justification of the project discussed here. All results gained during online survey are analyzed in this section. It also includes the system architecture and web based prototype.

4.1 Survey Analysis

A survey was conducted mainly to identify the awareness of recommender system and the users' attitude towards the recommender system in the e-business website. The survey was conducted online with 44 respondents. However, in total only 42 respondents are counts in certain questions because 2 of the respondents skipped certain questions. The results of the survey are very important in order to proof the idea of this project is acceptable and relevant by getting response from the respondents that mostly familiar with online businesses.

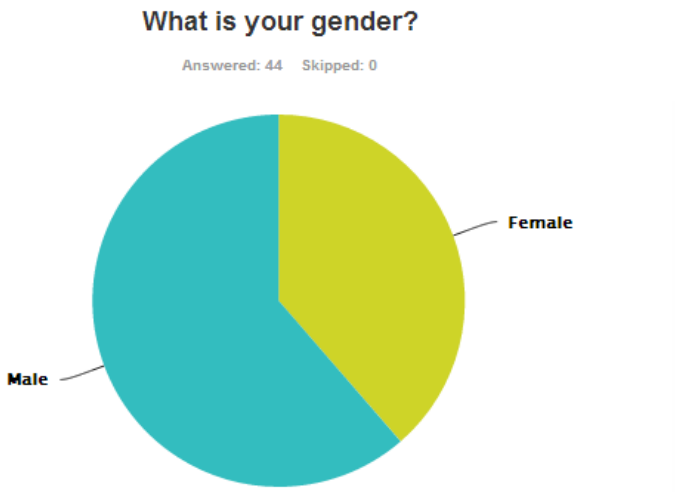
There are 10 close-ended questions being asked in the survey. There are three sections of questionnaires:

i. Section A: Demographic Data

In this section, the demographic data of the 44 respondents of the online survey were collected to obtaining specific information of the respondents. This is because the backgrounds of a person give effects in showing that person familiar with internet and online business or not. These data are then used in analyzing certain patterns and trends related to the project.

Three questions were asked, pertaining to the demographic data of the respondents. The questions and analysis are as followed:

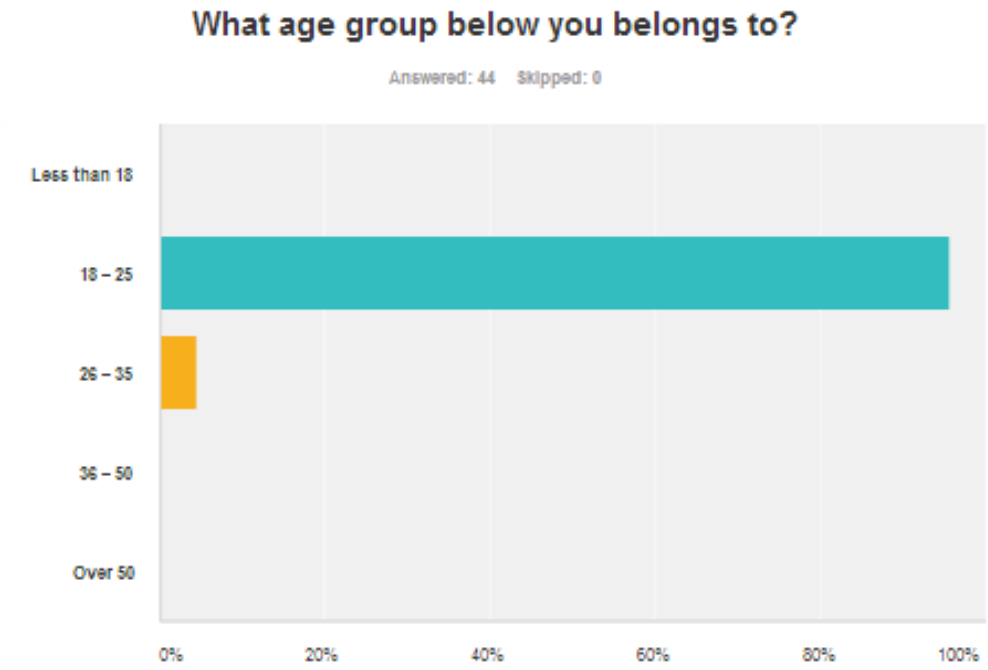
1.



Answer Choices	Responses	
Female	38.64%	17
Male	61.36%	27
Total		44

Analysis: Question 1 is referring to respondents' gender. 61% of the respondents are male and the remaining 39% are female. Hence, the majority of the respondents that respond to the survey questionnaires are males.

2.

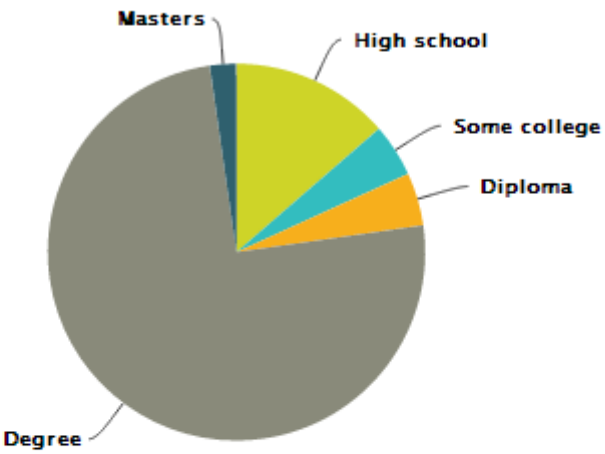


Answer Choices	Responses	
Less than 18	0%	0
18 – 25	95.45%	42
26 – 35	4.55%	2
36 – 50	0%	0
Over 50	0%	0
Total		44

Analysis: Question 2 is referring to respondents' age group. Majority of the respondents are 18 – 25 age groups (teenagers and young adults) which comprises 96% of the population sample. The remaining are from 26 – 35 age groups with 5%. Hence, it can be concluded that people from the 18 – 25 age groups are the major user of the Internet.

3. **What is your highest level of education?**

Answered: 44 Skipped: 0



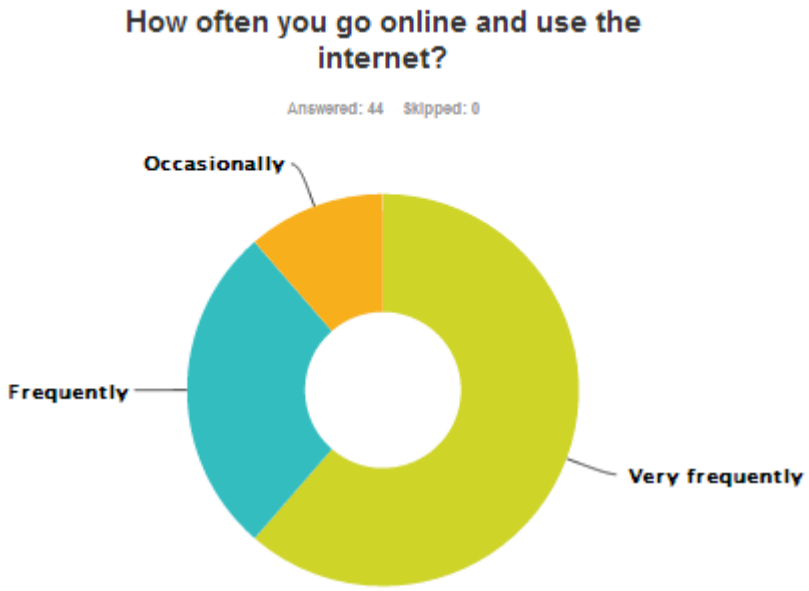
Answer Choices	Responses	
High school	13.64%	6
Some college	4.55%	2
Diploma	4.55%	2
Degree	75%	33
Masters	2.27%	1
PHD	0%	0
Total		44

Analysis: Question 3 is referring to respondents' current education level. 75% of the respondents hold a degree, only 13% with high school certificates, 5% each holding a diploma and some college and the other 2% with masters. It can be deduced that majority of the Internet users hold a degree.

ii. Section B: Awareness of Recommender System

In this section, the collected information is related with the familiarity with the Internet, online business and also the awareness towards recommender system in the e-business website. The questions and analysis are as followed:

4.



Answer Choices	Responses	
Very frequently	61.36%	27
Frequently	27.27%	12
Occasionally	11.36%	5
Rarely	0%	0
Very rarely	0%	0
Never	0%	0
Total		44

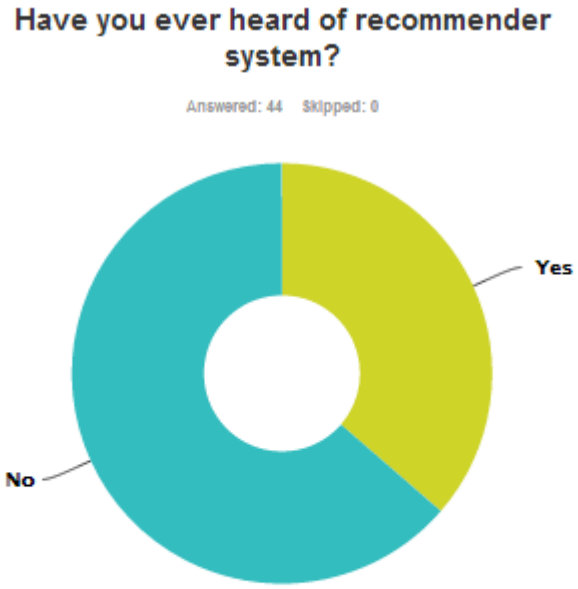
Analysis: Question 4 is asking the respondents on how often do they go online and use the Internet. 61% of the respondents said that they go online very frequently. While 28% said they frequently go online and 11% of them occasionally go online. Therefore, this reflects that most of the respondents are very frequently go online and using the Internet regularly in their daily life.

5.

Answer Choices	Responses	
Often	15.91%	7
Sometimes	38.64%	17
Seldom	31.82%	14
Never	13.64%	6
Total		44

Analysis: Question 5 is asking the respondents on how often do they purchase products or subscribe services online. 39% of the respondents said sometimes, 32% said seldom, 16% said often and the rest 13% said they never purchase products or subscribe services online. It can be deduced that most of the respondents just sometimes purchase products or subscribe services online.

6.



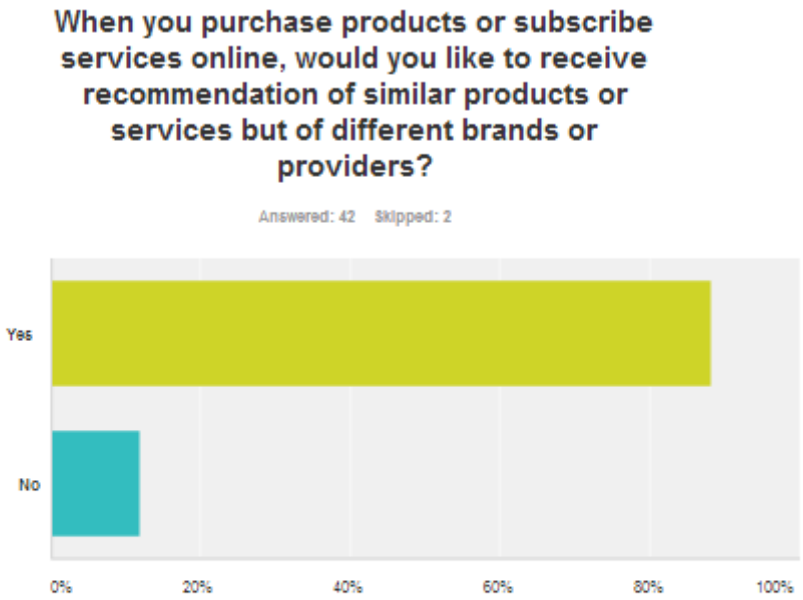
Answer Choices	Responses	
Yes	36.36%	16
No	63.64%	28
Total		44

Analysis: Question 6 is asking the respondents whether they have heard about recommender system. Majority (64%) said no and the remaining 36% of the respondents know about recommender system. Hence it can be concluded that majority of the respondents did not know about recommender system. It is maybe they know the system but they do not know its name, or they just do not exposed with the recommender system in the e-business website.

iii. Section C: Users Attitude towards Recommender System

In this section, the collected information is about the users' attitude towards recommender system. This section is very important in order to recognize the usability and effectiveness of the recommender system. The questions and analysis are as followed:

7.



Answer Choices	Responses	
Yes	88.10%	37
No	11.90%	5
Total		42

Analysis: Question 7 is asking either the respondents like to receive recommendation of similar products or services but of different brands or providers when they purchase products or subscribe services online. Majority (88%) said they would like to receive the recommendation and the remaining 12% do not like it. Hence, it can be concluded that majority of the Internet users like to receive recommendation of similar products or services but of different brands or providers when they purchase products or subscribe services online.

8.

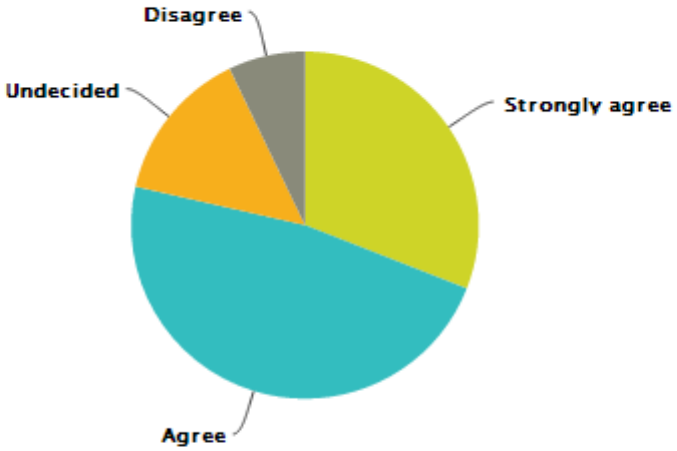
Answer Choices	Responses	
Yes	92.86%	39
No	7.14%	3
Total		42

Analysis: Question 8 is asking either the respondents like to receive recommendation of complimentary products or services but of different brands or providers when they purchase products or subscribe services online. Majority (93%) said they would like to receive the recommendation and the remaining 7% do not like it. Hence, it can be deduced that majority of the Internet users like to receive recommendation of complimentary products or services but of different brands or providers when they purchase products or subscribe services online.

9.

When you buy things online and the website wants you rate their products, you do not do it because it is annoying/troublesome.

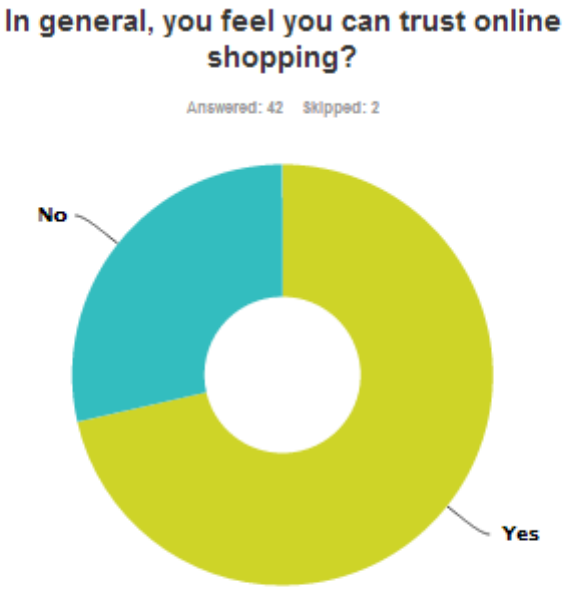
Answered: 42 Skipped: 2



Answer Choices	Responses	
Strongly agree	30.95%	13
Agree	47.62%	20
Undecided	14.29%	6
Disagree	7.14%	3
Strongly disagree	0%	0
Total		42

Analysis: Based on the statistics, 48% of the respondents agreed that they do not rate of the items that they buy online when the website wants them to do so because it is annoying or troublesome. 31% agreed with the statement, 14% is undecided and 7% is disagreed. Hence, it can be deduced that majority of the Internet users do not rate the items that they bought at e-business website because it is annoying or troublesome for them.

10.



Answer Choices	Responses	
Yes	71.43%	30
No	28.57%	12
Total		42

Analysis: Based on the graph above, 71% of the respondents said that they trust online shopping while the remaining 29% did not trust online shopping. Hence, it can be concluded that majority of the Internet user trust online shopping, and maybe partially of the respondent that trust online shopping used recommender system in determining which items that they want to buy.

4.2 System Flow Diagram

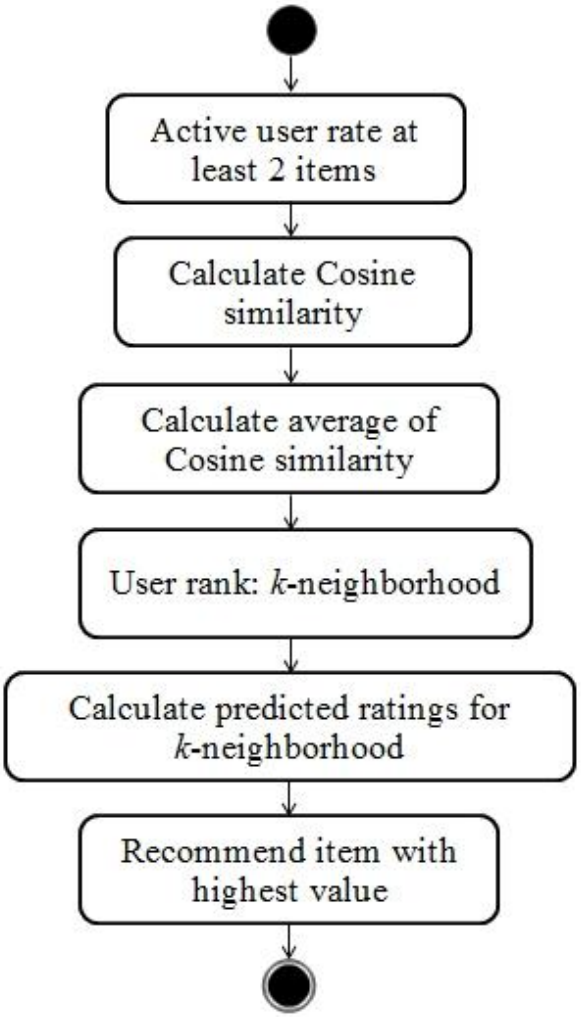


FIGURE 6: System Flow Diagram

Below is the process of the recommender system:

i. The active user need to rate at least 2 items.
ii. The system will calculate the Cosine similarity of the active user with each of the existed user in the database.
iii. The system will calculate the average of the Cosine similarity.

iv. The system will identify the *k*-neighborhood, the users that have Cosine similarity value above of the average, and then rank the user.

v. Calculated predicted ratings for the items in *k*-neighborhood by calculate the average of items that have not voted yet by the active user.

vi. The item that has the highest predicted ratings will be recommended to the active user.

4.3 System Prototype Design

In simulating the real environment of recommender system for e-business website, the author proposed to have a website using movies as domain and will recommend what movies that active user would like to download.

4.3.1 Database

Below are the screenshots of database of the system:

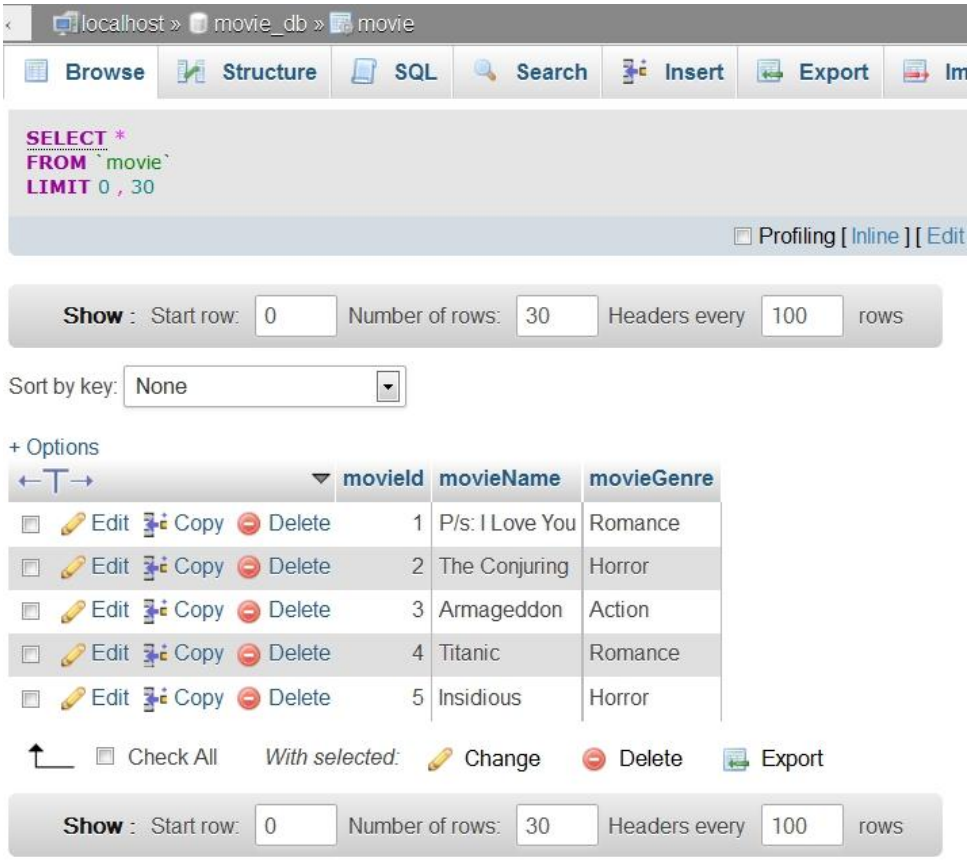


FIGURE 7: Movies Database

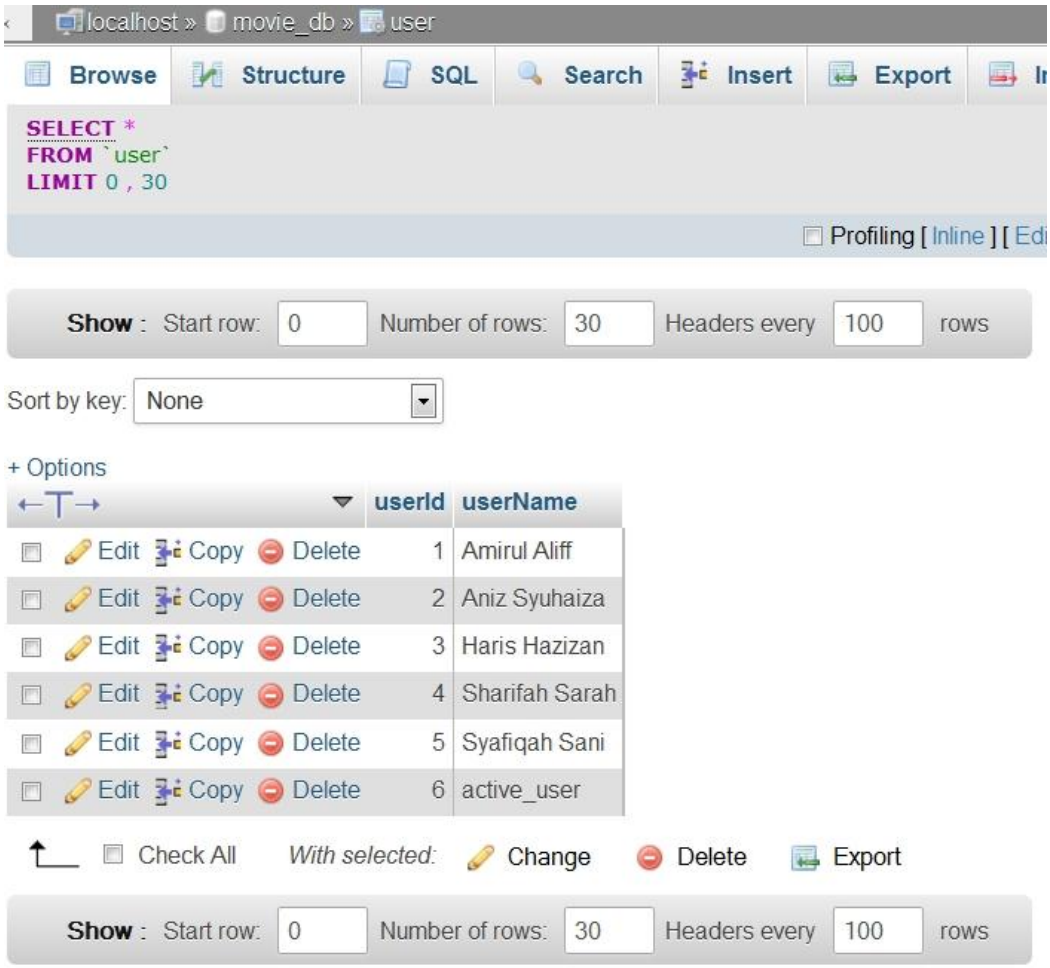


FIGURE 8: Users Database

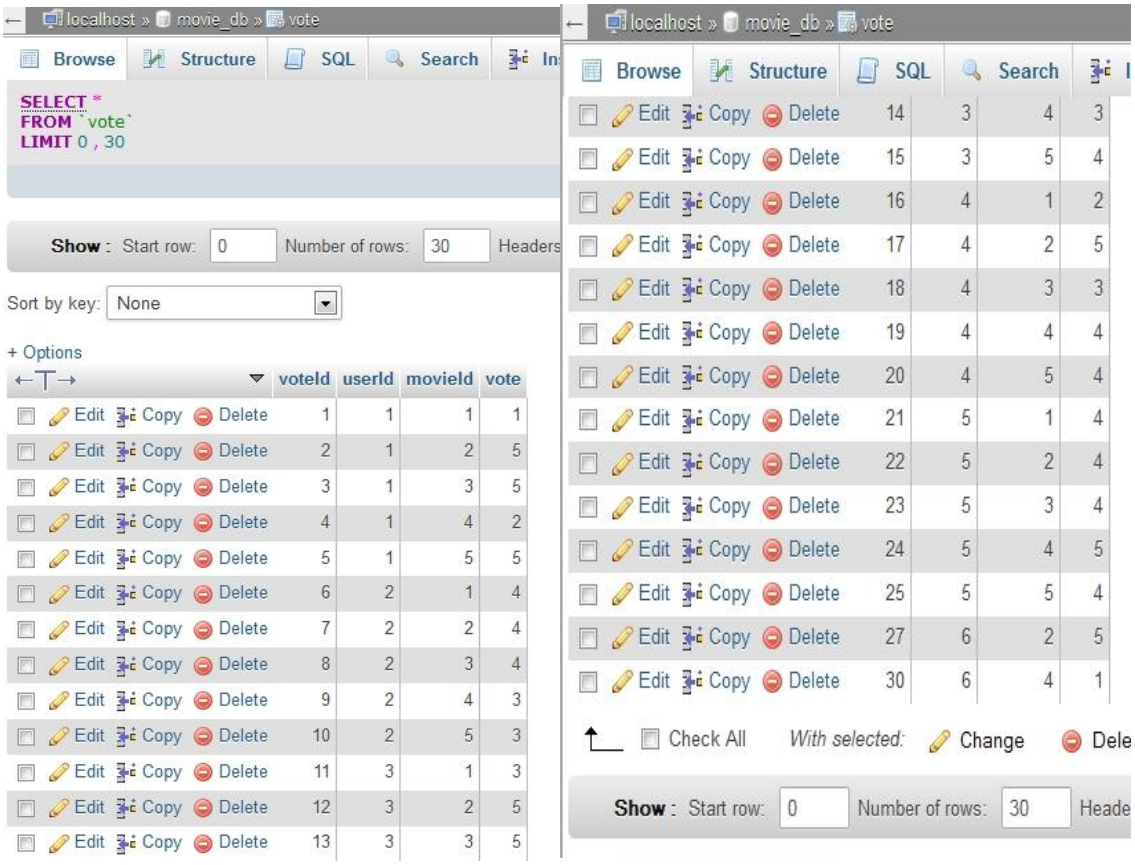


FIGURE 9: Ratings Database

.

	P/s: I Love You	The Conjuring	Armageddon	Titanic	Insidious
	Romance	Horror	Action	Romance	Horror
Amirul Aliff	1	5	5	2	5
Aniz Syuhaiza	4	4	4	3	3
Haris Hazizan	3	5	5	3	4
Sharifah Sarah	2	5	3	4	4
Syafiqah Sani	4	4	4	5	4
active user	?	5	?	1	?

FIGURE 10: Utility Matrix (Prototype)

Figure 9 shows the utility matrix of R with 5 users and 5 movies with ratings in the range 1 to 5. As discussed in the previous chapter, the system provides the recommendation for the active user. The Cosine Similarity algorithm and predicted ratings will be calculated using the above utility matrix.

4.3.2 Home Page

FIGURE 11: Home Page

The home page of the web page simulation will show a list of movies and its genre. This page will be automatically view as the active user because of the recommendation system will provide the recommendation movie for the active user.

4.3.3 Rated Movies Page

FIGURE 12: Rated Movies Page

This page will shows to the active user that the active user already voted the movie, so do not have to vote it again. For the simulator, the active user already rated two movies which are The Conjuring with 5 stars and Titanic with 1 star.

4.4.4 Recommended Movies Page

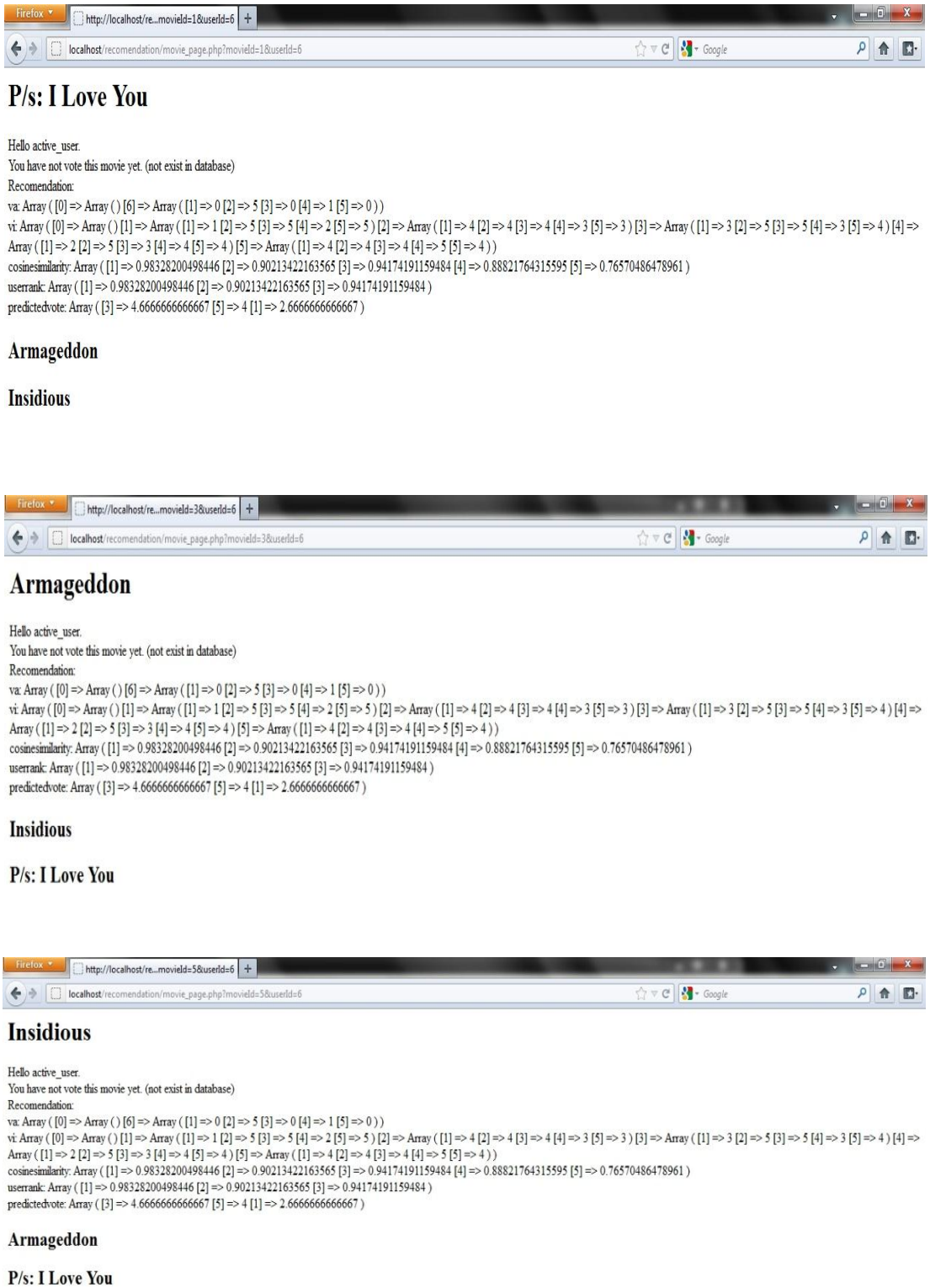


FIGURE 13: Recommended Movies Page

In the utility matrix that have been set earlier, there are two movies that already rated by the active user which are The Conjuring and also Titanic. Hence, there are three movies that still not rated by the active user which are P/s I Love you, Armageddon and Insidious. When the active user click one of this these three movies at the home page, the system will automatically calculated and predicted which movies will recommend for the active user.

4.4 System Evaluation

Below are the results from the calculation of the Cosine similarity algorithm and predicted ratings from the simulator:

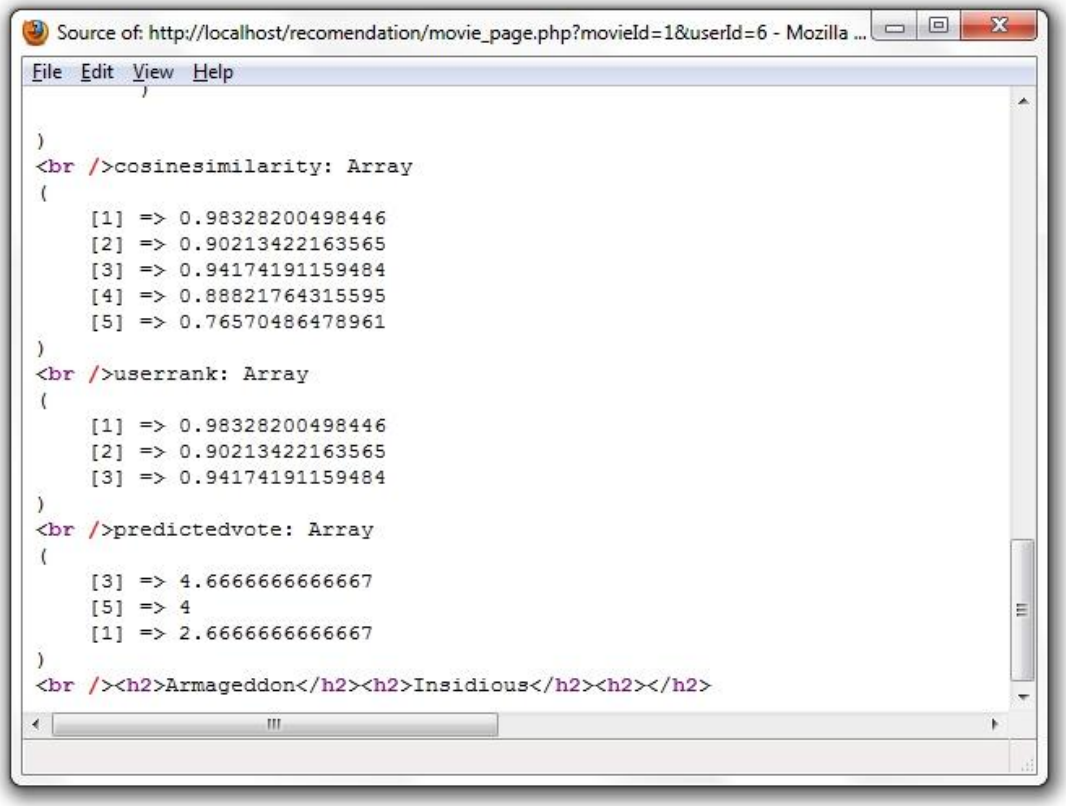


FIGURE 14: Results

4.4.1 Cosine Similarity

In Figure 13, the system listed out the value of Cosine similarity of the active user with the user 1 until user 5. From these values, the system will identify the *k*-neighborhood for the active user.

4.4.2 User Rank

The system calculated the average of the Cosine similarity and sorted the Cosine similarity value above of the average, which is the *k*-neighborhood. In this simulation, the k-neighborhood is user 1, user 2 and user 3. The user ranked based on the value of the Cosine similarity in the *k*-neighborhood, which shows active user have the most similar preference with user 1, followed by user 3 and then user 2.

4.4.3 Predicted Ratings

The predicted ratings for the movies in *k*-neighborhood by calculate the average of movies that have not voted yet by the active user which are P/s I Love you, Armageddon and Insidious. Based on the results of the simulation, the highest predicted rating is Armageddon with 4.6666666666667, followed by Insidious with 4 and P/s I Love you with 2.6666666666667. Based on the predicted ratings theory, the item that has the highest predicted ratings will be recommended to the active user. Hence, the movies that will recommend to the active user are Armageddon, followed by Insidious and the last one is P/s I Love You. From the results, we can see the pattern of the active user preferences is most similar to the user 1 and movie suggested by the system is match with both active user and user 1. Hence, the Cosine similarity algorithm and predicted ratings calculation proven to help in giving suggestion and give recommendation on what movie to download to the active user.

CHAPTER 5
CONCLUSION AND RECOMMENDATION

In conclusion, this research project is focusing to successfully develop recommender system for e-business website through the concept of Cosine similarity algorithm and calculation of the predicted ratings. Furthermore, this research project focuses on how to provide suggestion item for the active user or potential customer based on their preference. It is believed that the recommender system for e-business website will give benefits to both e-business website and also to the customers.

For recommendations, future area in this area needs to be done in order to refine the suitability of the system with content of e-business website. With that, the recommender system of this project can be embedded to a real e-business website to enhance the credibility and explore this system. Other than that, it would be great if future works will explore more features of recommender system more personalized not only based on the ratings compared to what is done in this project.

REFERENCES

1. Gary P. Schneider (2011). E-Business, 9th Edition. Course Technology, Learning Centre.
2. Uma Sekaran (1992). Research Methods for Business, 2nd Edition. John Wiley & Sons, Inc.
3. Cosly, D., Lam, SK., Albert, I., Konstan, JA., & Riedl, J. (2003). *Is Seeing Believing? How Recommender Interfaces Affect Users' Opinions.* University of Minnesota, Minnesota, MN 55455 USA.
4. Han, P., Xie, B., Yang, F., & Shen, R. (2004). *A scalable P2P Recommender System Based on Distributed Collaborative Filtering.* Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China.
5. Rashid, A., Karypis, G., & Riedl, J. (2007). *Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach.* Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN 55455 USA.
6. Ricci, F., Rokach, L., * Shapira, B. (2011). *Introduction to Recommender System Handbook.* Faculty of Computer Science, Free University of Bozen-Bolzano.
7. Schafer, JB., Konstan, J., & Riedl, J. (1999). *Recommender Systems in E-Commerce.* University of Minnesota, Minnesota, MN 55455 USA.
8. Nordan, NA., Abidin, A.I.Z., Mahmood, A.K., & Arshad, N.I. (2009). *Digital Social Networks: Examining the Knowledge Characteristics.* Department of Computer Information Science, Universiti Teknologi PETRONAS, Malaysia.
9. Kasunik, M. (2005). *Designing an Effective Survey*. Carnegie Mellon, Software Engineering Institute, Pittsburgh, PA.
10. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S. (2007). *Collaborative filtering recommender systems.* Springer berlin / Heidelberg.
11. Hahsler, M. (2011). *A Framework for Developing and Testing Recommendation Algorithms.* NSF Industry/University Cooperative Research Center for Net-Centric Software & Systems.

11. Collaborative Filtering. Retrieved July 30, 2013, from http://en.wikipedia.org/wiki/Collaborative_filtering.
12. Guide to Writing Survey Questions. Retrieved July 25, 2013, from http://www.mad.state.mn.us/survey-guide.
13. Survey Design Best Practice. Retrieved July 25, 2013, from http://help.surveymonkey.com/articles/en_US/kb/Design-Tips-How-to-create-and-administer-effective-surveys.
14. Questionnaire Development. Retrieved July 25, 2013, from http://www.esurveyspro.com/article-questionnaire-development.aspx

APPENDICES

Appendix A: Survey Questionnaires

RECOMMENDER SYSTEM FOR E-BUSINESS WEBSITE

The main objective of this project is to provide a recommender system for e-commerce website that give suggestion of items to the customers based on their preference. By using suitable algorithms, the suggested items will be predicted based on the review of the items and also the content of the items.

PART A: Demographic

1. What is your gender?
    - Male
    - Female

2. What age group below you belongs to?
    - Less than 18.
    - 18 – 25
    - 26 – 35
    - 36 – 50
    - Over 50

3. What is your highest level of education?
    - High school
    - Some college
    - Diploma
    - Degree
    - Masters
    - PHD

PART B: Awareness of Recommender System

4. How often you go online and use the internet?

- Very frequently
- Frequently
- Occasionally
- Rarely
- Very rarely
- Never

5. How often you purchase products or subscribe services online?

- Often
- Sometimes
- Seldom
- Never

6. Have you ever heard of recommender system?

- Yes
- No

PART C: Users Attitude towards Recommender System

7. When you purchase products or subscribe services online, would you like to receive recommendation of similar products or services but of different brands or providers?

- Yes
- No

If no, please state your reason.

8. When you purchase products or subscribe services online, would you like to receive recommendation of complimentary products or services of the product you purchase or the service you subscribe?

- Yes
- No

If no, please state your reason.

9. When you buy things online and the website wants you rate their products, you do not do it because it is annoying/troublesome.

- Strongly agree
- Agree
- Undecided
- Disagree
- Strongly disagree

10. In general, you feel you can trust online shopping?

- Yes
- No

If no, please state your reason.