Website Personalization Based on Demographic Data

By

Intan Najua binti Kamal Nasir

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

JULY 2005

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

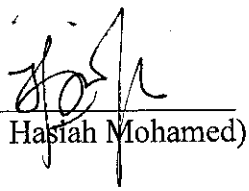
Website Personalization Based on Demographic Data

By

Intan Najua binti Kamal Nasir (3059)

A project dissertation submitted to the
Information System Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirements for the
Bachelor of Technology (Hons)
(Business Information System)

Approved by,

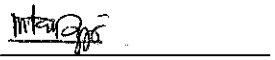


(Mrs. Hasiah Mohamed)

Universiti Teknologi PETRONAS
Tronoh, Perak
July 2005

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 as specified in the references and acknowledgements, and the original work contained herein have not been undertaken or done by unspecified sources or person.



INTAN NAJUA BINTI KAMAL NASIR

3059

ABSTRACT

This study focuses on websites personalization based on user's demographic data. The main demographic data that used in this study are age, gender, race and occupation. These data is obtained through user profiling technique conducted during the study. Analysis of the data gathered is done to find the relationship between the user's demographic data and their preferences for a website design. These data will be used as a guideline in order to develop a website that will fulfill the visitor's need. The topic chose was Obesity. HCI issues are considered as one of the important factors in this study which are effectiveness and satisfaction. The methodologies used are website personalization process, incremental model, combination of these two methods and Cascading Style Sheet (CSS) which discussed detail in Chapter 3. After that, we will be discussing the effectiveness and evaluation of the personalization website that have been built. Last but not least, there will be conclusion that present the result of evaluation of the websites made by the respondents.

ACKNOWLEDGEMENT

First of all, I would like to express my grateful and thankful to Al-Mighty God for giving me strength and time in developing the website as well as report. Besides that, I would like to thanks my Supervisor, Mrs. Hasiah Mohamed@Omar who is not give up on me when I encountered problem difficulties in order to develop the website. Her support and advice will not be forgotten. To Mr. Mohd Nor Ibrahim and Mr. Nordin Zakaria, our FYP coordinator who patiently deal with us during this 6 months to complete the project.

My thanks go to Miss Siti Sarah Harun whose has indirectly helping a lot in doing the project. My sincere thanks also go to all my colleagues and course mate who have been helping and supporting me through all the difficulties that I faced during the development of the website. Their kindness, willingness to share knowledge is the most valued and remembered.

Not forgotten to my beloved family who are far away from me. Their support and caring made me feel they are near and made me strong to overcome problems which encountered during these 6 months.

Millions of thank and gratitude to all who are involved in making this project successful. May Allah bless them. Thank you very much.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

With the increasing the popularity of the World Wide Web (WWW) imposes new methods of design and development of online information services. Most of the web structures are large and complicated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them. Therefore, there is the need for an important technology for understanding user's behaviors on the Web.

Personalization can simply be defined as a strategy, a marketing tool, and an art. Personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in the content delivery framework to manipulate what information presented to the users and how it is presented.

Proper implementation of personalization can bring focus to message provided and delivered in websites. For users, personalization can give them a more visitor-oriented web experience, relevant and quick inform. In business, personalization can be valuable because it drives desired business results such as increasing visitor response and promoting customer retention. Unfortunately, poor implementation of personalization has the potential to increase the complexity of the website interface, complicate the user experience and orphans the content.

User profiling is essential in order to personalize a website. A system should be able to distinguish between different users or groups of users. This can be achieved through user profiling process where information about the characteristics, preferences and activities of website's visitors is collected. These data includes demographic information such as name, age, gender, country of the visitors which will be used in personalizing websites.

This study focuses on implementation of personalization technology for website development based on demographic data. Age, gender, rage, race and occupation will be used as the main data in differentiating and guiding the website design. The website content is based on the topic "Obesity". This topic is chosen because it is relevant to all range of website's visitors despite their background or profiles.

1.2 Problem Statement

This study focuses on websites personalization based on user's demographic data. The main demographic data that will be used in this study are age, gender, race and occupation. These data is obtained through user profiling technique conducted during the study. One method that will be used in gathering the user's information is questionnaires. From this method, data containing user's demographic information and user's preferences and interest can be obtained.

Analysis of the data gathered is done to find the relationship between the user's demographic data and their preferences for a website design. These data will be used as a guideline in order to develop a website that will fulfill the visitor's need. Focus will be given on the effectiveness of the website and user's satisfaction when using the website. The content of the website will be made suitable for each group of users to improve the website effectiveness and to create user's retention of a website to improve user's satisfaction.

In addition, HCI issues are considered as one of the important factors in this study. Effectiveness and satisfaction are two principles of usability to be discussed in this study. These principles of usability will be highlighted in discussing the result. In this context, effectiveness refers to the effectiveness of web personalization techniques in developing website and increasing the user's web surfing experience. Meanwhile, satisfaction refers to user's satisfaction towards the website design.

1.3 Objectives

- To implement website personalization technique in developing a website with the user's demographic data.

  The data will be gathered from larger number of users so that designing the website layout that meet the user's preferences will become more accurate.

- To explore Hypertext Preprocessor (PHP) and other language or techniques in order to produce a better website based on the demographic data collected.

  The website development will be integrated with the usage of MySQL as well as other programming language to produce a more interactive and fulfilling the requirements.

- To study some HCI issues related to website personalization technology.

  Two principles in usability will be emphasized in discussing the result, which is effectiveness and satisfaction.

- To test and to evaluate the usage of personalization technology.

  Testing and evaluation of the website will be conducted to test the usability and effectiveness of this technology.

1.4 Scope of Study

For this study, the scope is based on the demographic data selected – age, gender, race, and occupation. Upon entering the website, visitors will have to provide information which is name, age, gender, race and occupation. The website layout design display will be based on the data provided by visitors. The users will be categorized according to specific profiles based on the average preferences captured from the study and research conducted. The website content presented will be almost the same, but different in terms of the highlighted topics. The web content will be about Obesity. There will be seven main topics, which are the following:

- **About Obesity**
  This section will cover about the brief history of Obesity in US scenario.
- **Contribution Factors**
  This page describes all the causes that lead to Obesity.
- **Consequences**
  This section provides information on the negative impact that come from Obesity.
- **Body Mass Index (BMI)**
  This section provides information on calculating the BMI for user who would like to measure how critical their Obesity is.
- **The Healthy Diet**
  This section will provide some tips and information on healthy diet that should be exercised by all on order to stay a healthy life.

As visitors enter the website, they will be categorized into specific profiles (e.g. Profile A: consists of users in range of 15-24 year old, gender male, race Malay and job status as student). Visitors with this profile will be viewing the same web design and topic highlights. Meanwhile, visitors with a different set of profile will have different web design and topic

highlights. Visitors will only have to register once. The data entered will be saved into a database. New visitors will have to fit all the required data upon visiting the website.

1.5 The relevancy of the Project

This project will focus on the enhancing the website and the research based on the further research and study conducted. However, the project emphasize given is more towards the research in generating a web that can be personalized by using demographic data. Focus will be given on finding the relationship between the demographic chosen and the website user interface and on how these data can reflect the users in order to produce a websites that suits them. This project will perform the study on larger number of survey than in previous project. Besides that, more elements will be added in the website such as images and animations. Furthermore, the web content, Obesity is relevant and important topic to inform and to educate all visitors despite their age, gender or occupation about this disease.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The tremendous growth in the number and the complexity of information resources and services on the Web has made the Web Personalization an indispensable tool for both Web-based organizations and for the end user. The ability of a site to engage the visitors at a deeper level, and to successfully guide them to useful and pertinent information, is now viewed as one of the key factors in the site's ultimate success. Web personalization can be described as any action that makes Web experience of a user customized to the user's taste or preferences. This chapter will elaborate each of the keywords in this project title, **Web Personalization based on Demographic Data** as well as Obesity.

2.2 Web Personalization

Principal elements of Web Personalization include modeling of Web objects (pages or products) and subjects (users or customers), categorization of objects and determination of the set of actions to be recommended for personalization. Web personalization can be defined as a technology whereby a website "follows" the user's preferences. This technology involves a process of gathering user-information during the interaction with the user. It has been widely used in such areas such as emails, E-commerce websites and search engines. However, personalization techniques are usually used in tailoring the information services to personal user needs. It is aimed to improve the user's experience of a service. Therefore, this technology is now considered important ingredient in developing a web application.

A more complex definition of Web personalization can be stated as any action that adapts the information and services provided by a website to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the user's navigational behavior and individual interests, in combination with the content and the structure of the website. The objective of a web personalization system is to provide users with the information they want or need, without expecting from them to ask for it explicitly [Eirinaki, M. and Vazirgiannis, M (2003)].

Personalization is sometimes referred to as customization. But it is necessary to stress the differences between customization and personalization. Customization is the process whereby the user sets preferences at a site such as selecting categories of information categories of information to be displayed on a search engine "my" page. In customization, the site can be adjusted to each user's preferences regarding its structure and presentation. Every time a registered user logs in, his/her customized home page is loaded. This process is performed either manually or semi automatically. Customization occurs when the user can configure an interface and creates a profile manually, adding and removing elements in the profile. The control of the look and/or content is explicit and is user-driven, i.e. the user is involved actively in the process and user has control [Bonettt,M (2001); Eirinaki, M and Vazirgiannis, M. (2003)].

On the other hand, personalization is referred to as the process whereby a program "follows" what a user does on a website and tries to "match" the user's behavior by providing the information related to what the user has previously done. Systems modification and concerning the content or even the structure of a website are performed dynamically. This can occur either actively (the user is aware of the parameters and can have a direct influence on the program, for instances, rating books), or passively (the user does not alter the program directly, but does so only through their behavior). Sometimes, it requires the user's active involvement such as fulfilling questionnaires or some user's profile. However, the user is seen as being passive, or at least somewhat less control. It is the website that monitors, analyses and react to behavior, for instance, content offered can be on tracking surfing decisions [Bonettt, M (2001); Eirinaki, M and Vaziegiannis, M(2003)].

B.Mobasher (2004) stated that Web personalization can be describe as any action that makes the Web experience of a user customized to the user's taste or preferences. The ability of a site to engage visitors at a deeper level, and to successfully guide them to useful and pertinent information, are now has become as one of the key factors in the site's ultimate success. The writer also stated the principle elements of Web personalization are include modeling of Web objects and subjects, categorization of objects and subjects, matching between and across objects and/or subjects, and determination of the set of actions to be recommended for personalization. The example of Web objects are pages or products meanwhile the example of subjects are users and customers of the website.

2.3 Web Personalization Techniques

There are three general groups of approaches and techniques used in Web personalization:

- **Manual decision rule systems**
  Manual decision rule systems allow web administrators to specify rules based on user demographics or static profiles (collected through a registration process). The rules are used to affect the content served to a particular user.
- **Content-based filtering agents**
  Content-based filtering system generally relies on personal profiles and the content similarity of Web documents to these profiles for generating recommendations.
- **Collaborative filtering systems**
  Collaborative filtering systems typically take explicit information in the form of user ratings or preferences, and through a correlation engine, return information that is predicted to closely match the user's preferences.

There are several drawbacks to content-based or rule based filtering techniques for personalization. The type of input is often a subjective description of the users by the users themselves, and thus the system performance degrades over time as the profile

ages. Furthermore, using context similarity alone may result in missing important "pragmatic" relationship among Web objects based on how they are accessed by users. Collaborative filtering techniques have it own drawback. The most important drawback is it lack of stability since it involve a very large set of data and this may led to unacceptable latency for providing recommendation. Another drawback is it emanates from the sparce nature of the dataset. As the number of items in the database increases, the density of each user record with respect to these items will decrease. This, in turn, will decrease the likelihood of significant overlap of visited or rated items among pairs of users resulting in less reliable computed correlations.

In many Web sites, however, it may be desirable to integrate the personalization actions throughout the site involving different types of objects, including navigational and content pages, as well as implicit product-oriented user events such as shopping cart changes, or product information request.

2.3.1 Web Usage Mining

Web usage mining [Srivasta at al., 2000] has been proposed as an underlying approach for Web personalization [Mobasher et al., 200a]. The goal of Web usage mining is to capture and model the behavior patterns and profiles of users interacting with a Web site. The discovered patterns are usually represented as collection of pages or items that are frequently accessed b groups of users with common needs or interest. Such patterns can be used to better understand behavioral characteristic of visitors or user segments, improve the organization and structure of the site, and create a personalized experience for visitors by providing the dynamic recommendations. The flexibility provided by Web usage mining can help enhance many of the approaches discussed above and remedy many of the shortcomings. In particular, Web usage mining techniques, such as clustering, association rule mining, and navigational pattern mining, which rely on offline pattern discovery from user transaction, can be used to improve the scalability of collaborative filtering when dealing with click stream and e-commerce data.

The goal of personalization based on Web usage mining is to recommend a set of objects to the current (active) user, possibly consisting of links, ads, text, product or services tailored to the user's perceived preferences as determined by the matching usage patterns. This task is accomplished by matching the active user session (possibly in conjunction with previously stored profiles for that user) with the usage patterns discovered through Web usage mining. The usage patterns used in this context aggregate usage profiles since they provide an aggregate representation of the common activities or interest of groups of users. This process is performed by the recommendation engine which is the online component of the personalization system. If the data collection procedures in the system include the capability to tracks users across visits, then the recommendation can represent a longer term view of user's potential interest based on the user's activity history within the site.

The overall process of Web personalization based on Web usage mining consists of three phases:

- **Data preparation and transformation**

  This phase transforms raw Web log files into transaction data that can be processed by data mining tasks. Besides that it also includes data integration from multiple sources, such as backend databases, application servers, and site content.

- **Pattern discovery**

  A variety of data mining techniques can be applied to this transaction data in the pattern discovery phase, such as clustering, association rule mining and sequential pattern discovery.

- **Recommendation**

  The result from pattern discovery will be transform into aggregate user profiles. The recommendation engine considers the active user session in conjunction with the discovered patterns to provide personalized content.

2.3.2 Web Mining based-on Novel Pattern Recognition

From other perspective, A.Massimiliano (2004) present the *Web mining* techniques for Web personalization based on a novel pattern recognition strategy which analyzes and classifies both static and dynamic features.

This novel *Web usage mining* strategy consist of two phases: in the first one a pattern analysis and classification is performed by means of unsupervised clustering algorithm, using the registration information provided by the users. The second phase, reclassification is repeated until a suitable convergence is reached. Reclassification is used to overcome the inaccuracy of the registration information and it is accomplished by the log analysis and content management modules based on user's navigational behavior.

The used of unsupervised clustering procedure for portioning the feature space built upon the user-provided data into a certain number of clusters (each one representing a class) that group together users appearing to be similar. In order to choose the optimal number of clusters, the generalization capability is maximized. This is where the use of Autoclass takes part. Autoclass is a fuzzy unsupervised clustering algorithm based on the Bayesian theory. Each cluster is described through a likelihood function depending on some parameters. Given the numbers of classes, the Autoclass C Search module estimates such parameters on the training data and finds the partition of the feature space that maximizes the log-likelihood value. Once the optimal number of cluster has been chosen, the classification is performed by the prediction module ofAutoclass C. By using the Bayesian rule and the likelihood function of each class, it attributes a user to that class which exhibits the maximum a posteriori probability. If a new user registers to the website, it is classified according to the same scheme. Eventually, if a user explicitly changes the data in its registration form, it is classified again using the Autoclass C prediction module.

The reclassification phase is based on the interaction of each user with the web site. Assuming the interaction can be performed in three different ways:

- Queries containing the same keywords
- Searches among directory
- Navigation of some pages.

All the materials on the website is managed by the content management module of the system, which associates each resource (a keyword, a directory, a news headline or an article) to a specific content category. On the other hand the log analysis module records all the activities of the users. In order to use this information for reclassifying users we need to attribute each category to a specific user class. This class can be accomplished by considering the first classification performed by Autoclass C and counting number N*i* of times in which the users of the i-class requested resources belonging to a specific category, over a time interval T, each category is then attributed to the class that maximizes N*i*. This way of classifying the content categories can suffer the inaccuracy of the first classification. However if the time interval T is wide enough and the percentage of correctly classified user is acceptable, also the classification of the categories can be considered reliable. Now, a reclassification can be performed, by considering the resources that each user requested in a predefined time interval (reclassification period). If the majority of the requested contents belong to a class different from the initial one, the user is reclassification process will lead to convergence if, after a suitable number of reclassifies users goes to zero.

2.3.3 Integrating Domain Knowledge with Web Usage Mining

Kathy Dai and B.Mobasher (2002), introduced a road map to more effective Web personalization integrating domain knowledge with Web usage mining. Personalization based on Web usage mining can enhance the effectiveness and scalability of collaborative filtering. However, without semantic knowledge about the underlying domain, such systems cannot recommend different types of complex objects based in their underlying properties and attributes. There are two general approaches to integrate semantic knowledge extracted from the content features of pages into the usage-based personalization process. Next, there are a general framework of integrating domain ontology with Web Usage Mining and Personalization.

At a conceptual level, there may be many different kinds of objects within a given site that are accessible to users. At the physical level, this object may be represented by one or more Web pages. For instance, the movie site mentioned earlier may contain pages related to the movies, actors, directors, studios, etc. conceptually, Each of this entities represents a different type of semantic object. During a visit to this site, a user may access several of these objects together during a session. In contras to content features, ontological representation of domain knowledge contained in the site makes it possible to have a uniform architecture to model such objects, their properties and their relationships.

The ontology preprocessing phase takes as input domain information (such as database schema and metadata) as well as Web pages, and generates the site ontology. For simple Web sites, ontologies can be easily designed manually or derived semi-automatically from the site content. However, it is more desirable to have automatic ontology acquisition methods for large web site, especially in web sites with dynamically generated web pages. E-commerce web sites, for instance, usually have well-structured web content, including predefined metadata or database schema. Therefore it is easier to built automatic ontology abstraction mechanisms that are sites-specific.

There have been a number of efforts dealing with the ontology learning problem. A wide of information, such as thesauri, content features, and database schema can help to identify ontologies. In the notion of "Semantic Web Mining" was introduced, including a framework for the extraction of a concept hierarchy and the application of data mining techniques to find frequently occurring combinations of concepts.

Domain ontologies can be incorporated into usage preprocessing to generate semantic user transactions (pre-mining), or they can be integrated into pattern discovery phase to generate semantic usage patterns. Here the focus id on the latter approach.

Given a discovered usage profile (a set of page view-weight pairs), as describe earlier, we can transform it into domain-level aggregate representation of the underlying objects. To distinguish between the representations we call the original discovered pattern as "item-level" usage profile, and we call the new profile based on the domain ontology a "domain-level" aggregate profile. The item-level profile is first represented as a weighted set of objects. The profile represent a set of objects accessed together frequently by a group of users (as determined through Web usage mining). Objects, in the usage profile, that belong to the same class are combined to form aggregated pseudo object belonging to that class. An important benefit of aggregation is that the pattern volume is significantly reduces, thus relieving the computation burden for the recommendation engine. The goal is to create an aggregate representation of this weighted set of objects to characterize the common interest of the user segment captured by the usage profile at the domain level.

2.4 Demographic Data

Demography is referred to as a study of statistic such as birth, deaths, disease, population etc. In order to show the state of community and demographic data referred as static data such as race, age, gender, occupation, income, address and many more. However, for this study context, the focus will be on four main demographic data, which are gender, age, race and occupation. These demographic factors will be used as differentiating measures to differentiate the user's preferences in order to design the website.

The relation of these four factors will be used as the guideline to design the website. These factors can reflect user's characteristics and as a result, the characteristics will determine the user interface design. A study by Plocher T. A., Garg C, and Chestnut J. (1999), showed how culture as one of demographic data determined user interface preferences. The study also highlighted, how factors such as language and family and societal structure affected the user interface preferences.

Shih, H. M. and Goonetileke, M. J. (1997) had conducted a study focused on the effects of language flow (for instance the intrinsic reading and writing direction of writing) on preferences for menu orientation. He found that Chinese performed signicantly better with horizontal-oriented menus than with vertical menus. Their explanation was that the horizontal orientation was better able to "break" a subject's natural reading flow in Chinese (top to bottom, right to left) than vertical menus, and hence was a more natural and effective search direction.

Meanwhile, Choong Y.Y. (1996), in his study explored the implications of the ironic nature of Chinese language for user interface design. Based on the theoretical work by Hossain R. (1986) and Liu I. M. (1986), Choong hyphotesized that, due to the pictorial nature of Chinese characters, Chinese users would prefer graphical over alphanumerical presentations in icon-driven. In contras, English-speakers should prefer more textual, alphanumerical presentations.

Obviously, personalization can be very helpful in order for users to emphasize better web experience to the users. But there are also many issues such as privacy that need to be considered when user provides their personal information. In March 2000, the Personalization Consortium conducted a survey of 4,500 web-user's opinions on personalization and online privacy. The result showed privacy issues are important to users, but do not outweight benefits that come from sharing personal information. The survey indicated that only 15% of web-users would be unwilling to provide personal information to web-marketers if that information improved their online service experience. 51% of respondents said they would share personal information in exchange for better service, while 33% had no opinion.

User involvement is important to get the information needs and to implement the personalization technology. Thorough analysis and understanding user's demographic data is important to enable web developer to predict the website design and meet their preferences. Besides, by understanding these data, HCI issue such as privacy or accessibility can be tailored to fit the users.

2.5 Related Work in Personalization

K.Harris (2000) defined personalization as the ability for a web user to customize the content and layout of their own portal web page. It is one of the most popular ways of increasing traffic at portal sites today, and helps to endure return customers. A personalized page brings news and information to us. It allows us to see what we want, when we want. With personalized page, we can check our stocks, our weather and our sports teams. With information scattered all over the web, the ability to create our own on-line, real-time interactive newspaper is quite valuable.

Personalization, which allows a web user to choose the content and layout of their own portal web page, is one of the most popular ways of increasing traffic at web sites. But, to be successful, it must be simple and it must be intuitive. This paper presents common personalization features used by top portals and review the design of the interfaces of three top portals: My Excite, My Yahoo and MSN. This paper provides examples of good and bad design techniques used in the portal sites, and gives tips on how to design usable personalization features.

The review is based on the following areas:

Visibility and Readability

Make primary links and actions visible and obvious. Design pages for scanning, using highlighted text, bulleted lists and short sentences.

Simplicity

Keep frequent or critical tasks short and simple. Terminology should be based on the user's language. The less the better.

Performance

Design pages to download quickly, for more than half of the user population still accesses the Internet at 33.6kb or lower.

Navigation and Organization

Provide clear methods of continuing, canceling or going back, and going on every page. Provide effective page titling to keep the user informed of location. Organize pages so that related information is grouped together and easily accessed.

Consistency

Similar task should be performed similarly. Reduce the need for users to learn multiple behaviors and navigation paths.

Feedback

When there is a problem, a message should tell the user exactly what's wrong and how to fix it in language they understand.

Tolerance

Provide forgiving systems that minimize the cost of user mistakes and allow users to undo their actions.

K.Harris also touched about good web design. The ability to provide personalized web content is vital to the success of the web. Personalization designers should strive to make these features simple and efficient. Two things to remember during design is to keep graphic to a minimum, and to design pages for 800 x 600 display resolution. For most users, the web is still a very slow experience, for more than half the user population still accesses the Internet at 33.6kb or lower. People are less willing to explore and navigate between multiple pages in this environment. He also advised to provide simple and fast features.

2.6 Health Issue

The usage of web personalization technology has increased into many other fields related to consumers. Besides E-Commerce, E-Health growth has also increased. Some Internet health services that are available nowadays are:

- E-Commerce
- Telemedicine consultants
- Prescriptions
- Provider credentialing
- Communication of clinical data to providers
- Forums to link to other providers, consumers/healthcare center
- Consumer education
- Patient satisfaction surveys
- Providing clinical pathways, practice guidelines
- Personal medical records

These services give a lot of benefits to users such as access to health information, connectivity to personal health community, personalization/health management, transactions/E-Commerce and services/telemedicine.

In this project, the topic chosen, Obesity can be categorized as one of the E-Health services. The related areas of Internet services are consumer education, providing clinical pathways, practice guidelines and links to other healthcare center. This title is suitable to educate Malaysian about the obesity since the awareness of this disease is not widely known among them.

Obesity is an abnormal accumulation of body fat, usually 20 percent or more over an individual's ideal body weight. Obesity is associated with an increased risk of illness, disability, and death.

During the past 20 years, obesity among adults has risen significantly in the United States. The latest data from the National Center for Health Statistics show that 30 percent of U.S. adults 20 years of age and older—over 60 million people—are obese.
This increase is not limited to adults. The percentage of young people who are overweight has more than tripled since 1980. Among children and teens aged 6–19 years, 16 percent (over 9 million young people) are considered overweight.

These increasing rates raise concern because of their implications for Americans' health. Being overweight or obese increases the risk of many diseases and health conditions, including the following:

- Hypertension
- Dyslipidemia (for example, high total cholesterol or high levels of triglycerides)

Type 2 diabetes

- Coronary heart disease
- Stroke
- Gallbladder disease
- Osteoarthritis
- Sleep apnea and respiratory problems
- Some cancers (endometrial, breast, and colon)

Although one of the national health objectives for the year 2010 is to reduce the prevalence of obesity among adults to less than 15%, current data indicate that the situation is worsening rather than improving.

From these facts, we can see the importance of Obesity website especially for Malaysian. Based on the researches that have been done on the Internet, most of the Obesity websites are from outside and the website that inform about Obesity's information is too little. So, by implementing personalization like this, it is a hope that consumer will visit this kind of website more frequently and in returns they will gain all the information that is meant for

them easily. As a result, they will be able to make a better decision about their own health and help to improve them for the future.

CHAPTER 3
METHODOLOGY

3.1 Procedure Identification

For completion of this project, a combination of two main methods is selected as a guideline:

- Website Personalization Process
- Incremental Model

3.1.1 Website Personalization Process

The steps of a Web personalization process include:

- Collection of data
- Modeling and categorization of these data (preprocessing phase)
- Analysis of the collected data
- Determination of the actions that should be performed (design)

The first step is the collection of data from user. The data collection is done through user profiling technique, which will result in the creation of an information base that contains the preferences, characteristics and activities of the user.

This technique can be done either explicitly or implicitly. Explicit collection of user profile data is performed through the use of online registration forms, questionnaires and the like, resulting in static user profile's data. Static user profile's data are information that is never or rarely altered, for instance, user's demographic information. Meanwhile, implicit collection of user profile data varies from the use of

cookies or similar technologies, resulting in dynamic user's profile. Dynamic user's profile is information that changes frequently such as user's navigational behavior or user's preferences.

In this step, user involvement is important in order to get the required information. Level of user involvement is also an important aspect in personalization. This is because:

- Too much explicit user involvement up front usually turns users away
- A mix of explicit and implicit data collection over time support live personalization and allows users to build up a sense of trust before they commit more sensitive profile information
- Users can set content profiles to affect the user's personalization results.

The second step, modeling and categorization of these data (preprocessing phase) involve classifying of collected data. Both static and dynamic user profile's data that are collected can be classified into four categories; user profile, usage, content and structure.

The third step involved the analysis of information gathered. The analysis of the colleted data is important step as the data will determined the design and layout of the personalize website. The data are organized and depicted into table, graphs and charts to make it easier to understand.

The last step is determination of the actions that should be performed. This step involves determining the proceeding task that should be on based on the analysis of the collected data. In this step, all the data gathered are used to decide weather to proceed with other crucial step steps or to repeat certain steps in the process in the case of error or inaccurate data. Figure 3.1 show the flow of website personalization process.

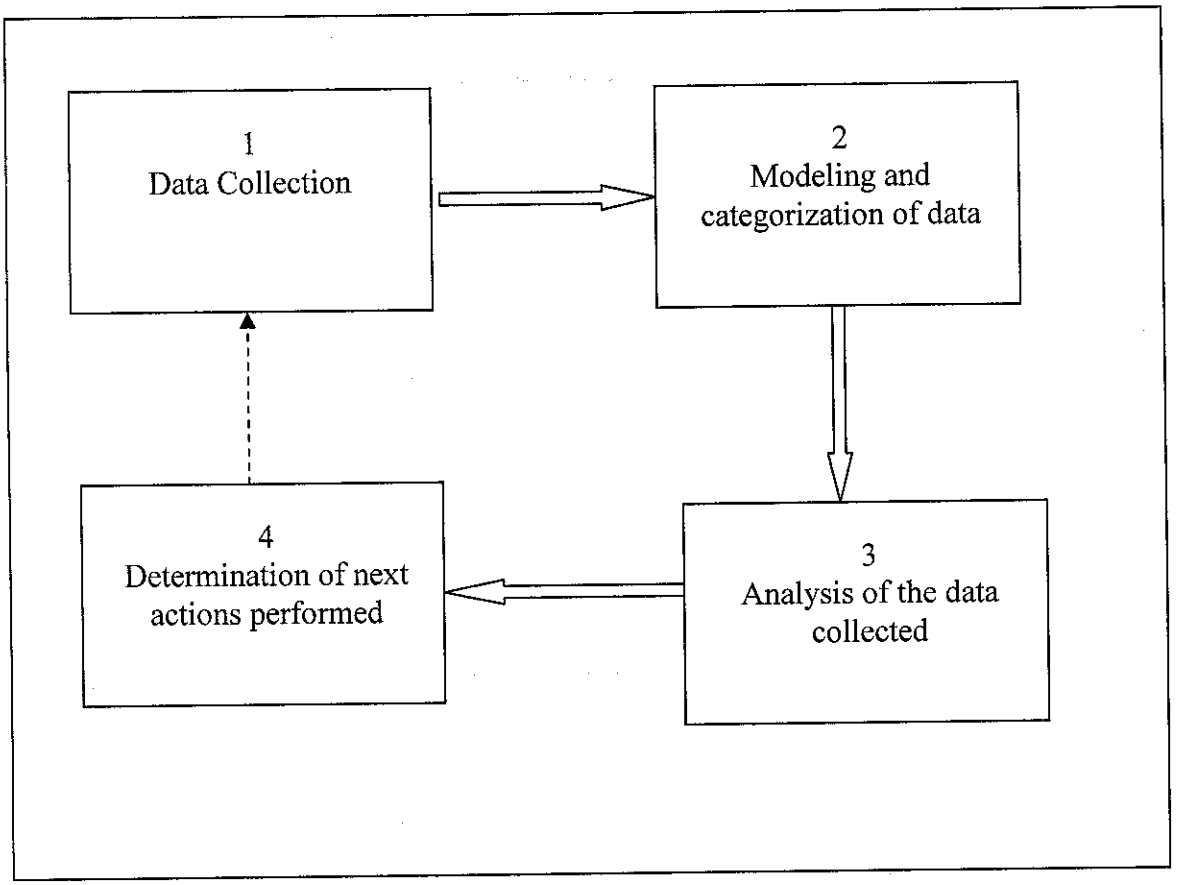


Figure 3.1: Website personalization process

3.1.2 Incremental Model

Incremental model will be used as a guideline for the website development. With this model, project is constructed step by step. The project is divided into series of builds. Each builds consists a set of module that provide functionality. The model is chosen because of its flexibility and its advantages such as:

- Products can be delivered in a portion base on its operational quality in short time
- Reduce the effect of introducing new system to users, which can give less traumatic effect to users
- Requires small capital outlay
- The process of construction can be stop at the end of any build
- Used open architecture, which is good for maintenance.

This model involves four phases; analysis, design, code and test. All these steps will be repeated in each build until retirement of the process where the product is completed. However, this model can be risky as product is developed piece by piece. There is a risk that pieces may not fir each other. Figure 3.2 shows the flow of incremental model:

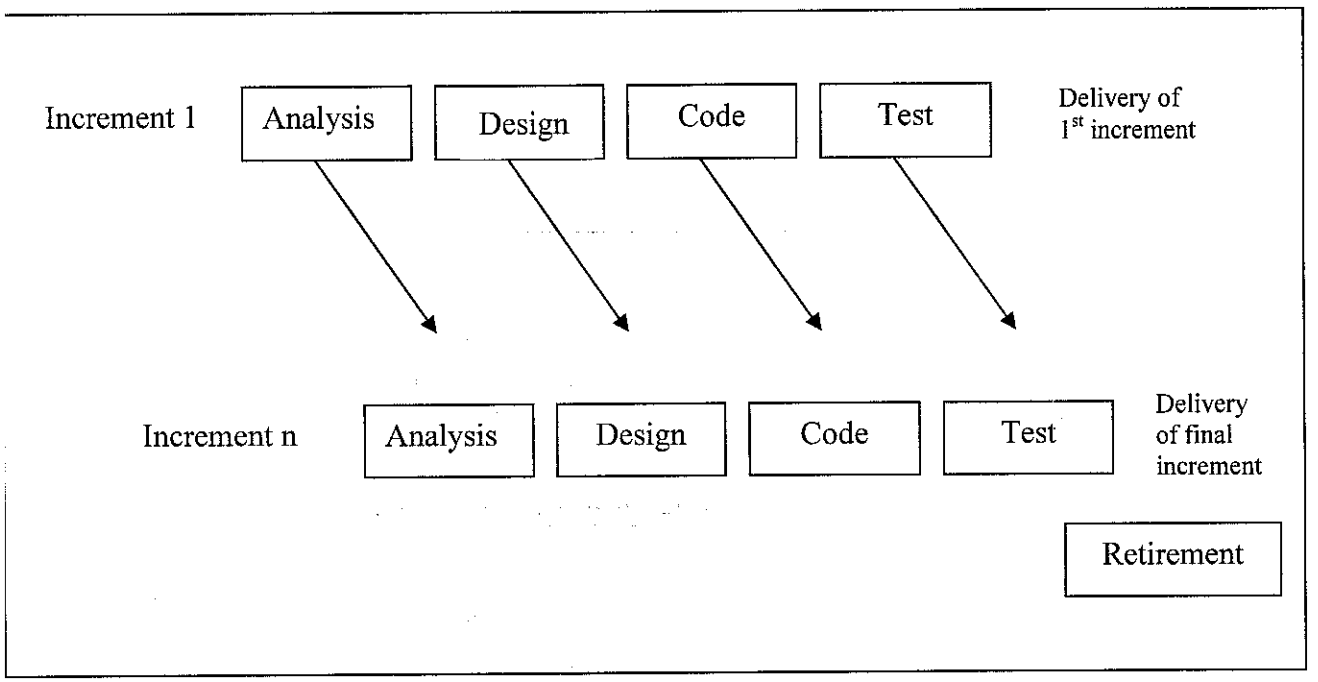


Figure 3.2: Incremental Model

3.1.3 Combination of methods

From the two methods mentioned earlier, the website personalization process and incremental model, a combination method is produced to be used a methodology outline for this project. The process flow of the steps involved in show in figure 3.3:

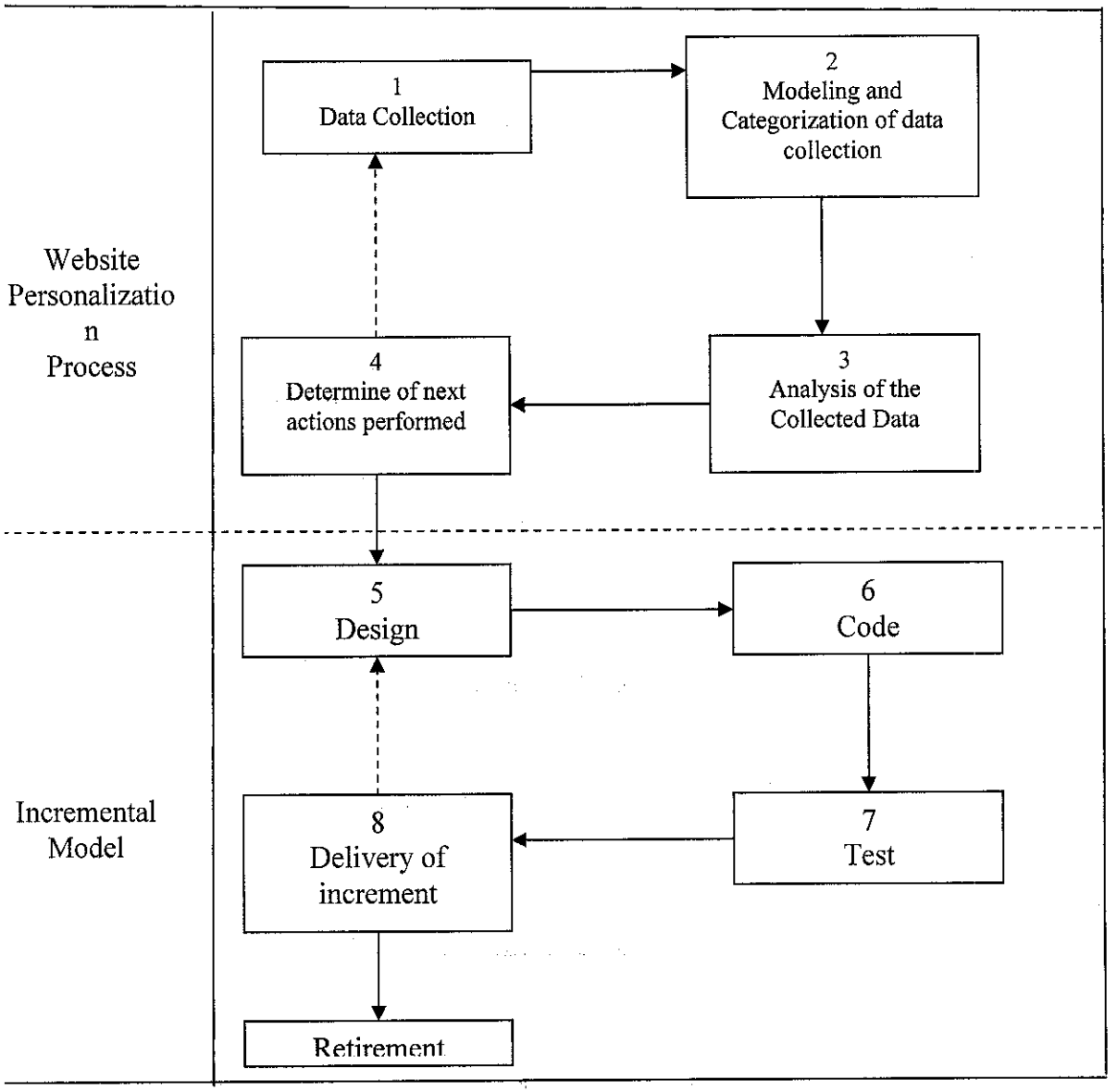


Figure 3.3: Steps in Combination Method

This combination method consists of eight steps. The first step from the incremental model is the analysis step is eliminated since the process of gathering and analysis of data is done in steps 1 to 4 in this combination model. Below are details for all the steps conducted.

3.2 Data Collection

In the first step, collection of data from users is done to get user's profile data such as user's preferences, user's profile, and user's surfing habits. The data is collected through the explicit data collection method, using questionnaires. The questionnaires are designed based on four focus area:

- Section 1 – Personal Details
- Section 2 – Internet Usage and Knowledge
- Section 3 – Website Content, Layout and Design.

In the first section, focus on the user's personal details. The main personal details that are important to capture are the respondent age, race, gender and job position level (occupation). These are the factors that will be used in determining the layout and design of the website that will be developed.

In the second section, the focus is on user's Internet knowledge and usage. From this section, information that can be captured is the user's surfing behavior such as how many hours they spend to surf the Internet per day and how many days they spend to surf the Internet per week. This information can determine whether the user is familiar with Internet and web facilities or not.

Section three is important as this section focus on the user's preferences on website content, layout and design. User's preferences of content, layout and design factors such as text color, background color type of text, picture position and content preferences are asked in this section. From this section, relationship between user's demographic data and user's preferences on website design and layout can be determined based on user's answer. This information will be used in designing the website.

The survey area for this project is based in UTP. The questionnaires are distributed among UTP population including UTP students and UTP staffs. Number of targeted respondent is based on the number of profile and number of sample per profile. For the purpose of this project, total number of sample targeted is 270 respondents consist of 9 profiles.

3.2.1 Modeling and Categorization of Data

The second step involved modeling and categorization of these data (preprocessing phase). Data that are collected from the questionnaires are classified in four categories:

- User profile – data represent personal information about the users contained demographic information such as name, age, gender, race and job position.
- Usage – data represent user's websites or Internet usage such as how often does the user used the Internet or how often does the user visit a website. User's knowledge and perception on the Internet usage is also included.
- Content – data represent data that user can 'see'. Including text style, background color, content color and display of the content.
- Structure – data represent the way content is organized. These includes the website layout, picture position, and frame design.

3.2.2 Analysis of Collected Data

The third step is analysis of collected data. Data from the questionnaires will be analyze and put into table, chart or graph. In this section, focus will be more the user's preferences for website layout, design and content. Data collected and result will be discussed in the next section – System Design and Development

3.2.3 Determination of Next Action to Perform

The fourth step is the determination of next action performed. This step involves making decision on what to do next base on available data. For this project, base on the data gathered, website layout and design can be determined. The design and layout will be discussed in the next section – System Design and Development.

3.2.4 Design

The next step involve in designing the website user interface and the database. The design and layout of the website will be based on the result from the previous step. At this stage, all four steps from website personalization process are completed and the following steps are from the incremental model. Based on incremental model, product is divided into build. For this project, each build id referred to one profile. Completion of each build will mean completion of the website design for one profile. In this context, there will be nine increments since there are nine profiles.

3.2.5 Code

The sixth step will be the crucial part in the development of the website. This step will involve in coding of the website functionality. The website functionality will be based on the coding which will be determined by the developer intention.

3.2.6 Test

Based on the incremental model, product can be delivered in a portion base on its operational quality. After completion of the website coding for each build, testing will be conducted to test the functionality. At this phase, the test will only focus on the functionality such ad detecting any error in terms of the user's registration, website design and layout, and the flow of the website content. This testing does not require user

involvement. User involvement will be during the testing conducted for delivery of increment/ build in the evaluation stage.

3.2.7 Delivery of Increment

After the testing is done, the build will be delivered to user. At this stage, user will involve in the testing process to evaluate the functionality of the website and the website effectiveness as well as to see their satisfaction towards the website. Based on the user's responds from the evaluation, it will be determined whether there is enhancement or amendment need to be done or to proceed with the next build. If enhancement or amendment needs to be done, repetition of increments step occurred. Each build will undergo the same steps until the retirement of each build. When all the builds delivered have satisfied the users, combination of all builds will be done and the users will evaluate the complete product again. If the product is success and the next step will be retirement of the project. With the project retirement, all the steps in the combination method are completed.

3.3 Processed Workflow

There are a few flows that have been determined. The process workflow will determine the general navigation of the website as a whole. It can be easily depicted as in the diagram below.

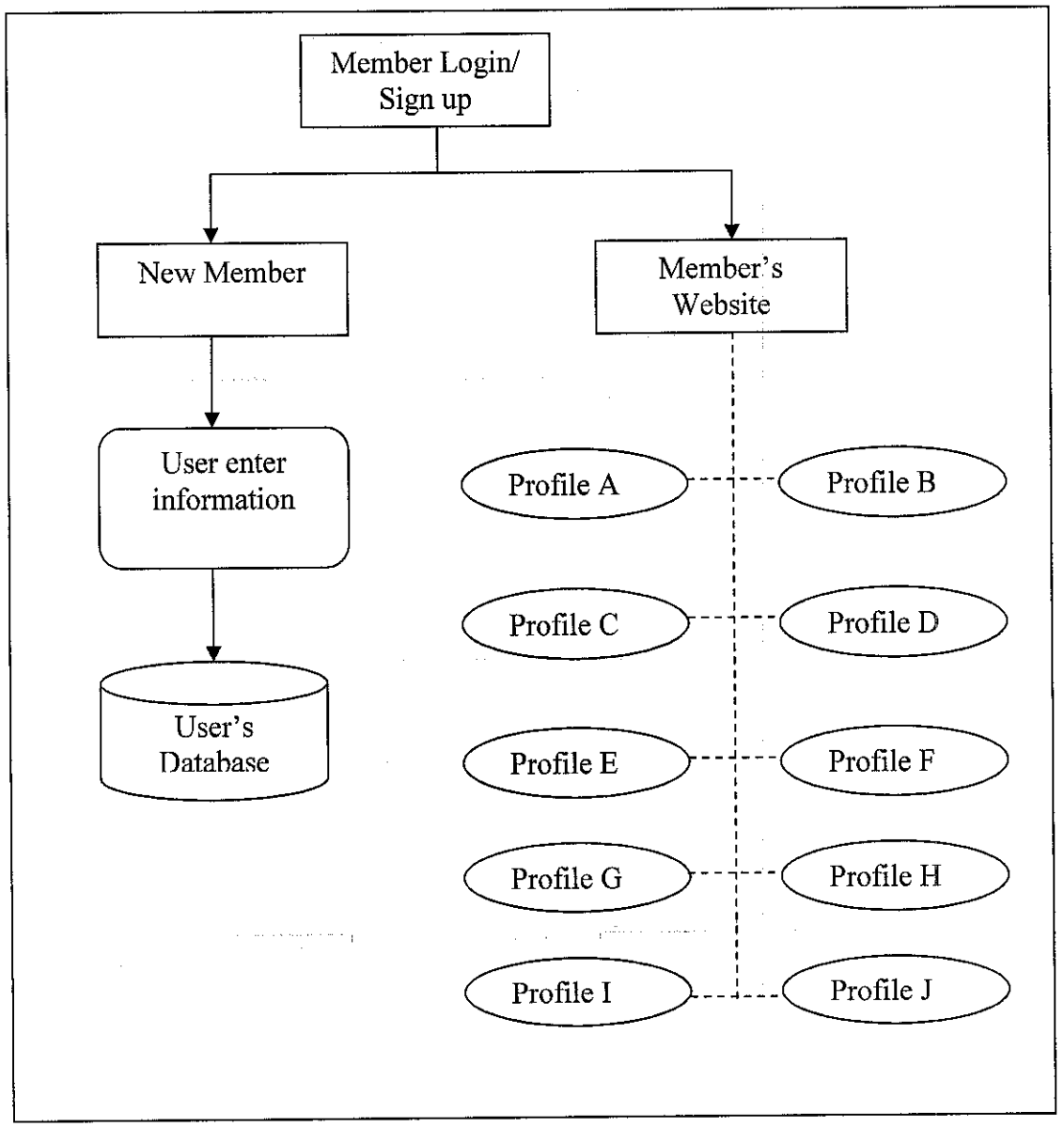


Figure 3.4: Process Workflow

3.4 Cascading Style Sheets

In order to format and to present information, there is a W3C technology called Cascading Style Sheet (CSS) format that allows document authors to specify the presentation of elements on a Web page (For instance fonts, spacing, and margins) separately from the structure of the documents (For example section headers, body text, and links). This separation of structure from presentation simplifies maintaining and modifying a document's layout. In other words, it tells you what you can say and how you can say it.

CSS is used by both the authors and readers of web pages to define colors, fonts, layout, and other aspects of document presentation. It is designed primarily to enable the separation of document structure (written in HTML or a similar markup language) from document presentation (written in CSS). This separation provides a number of benefits, including improved content accessibility, greater flexibility and control in the specification of presentational characteristics, and reduced complexity of the structural content. CSS is also capable of controlling the document's style separately in alternative rendering methods, such as on-screen, in print, by voice (when read out by a speech-based browser or screen reader) and on Braille-based, tactile devices.

Advantages of using CSS include:

- Presentation information for an entire website or collection of pages resides in one place, and can be updated quickly and easily—that is, if a style sheet is imported.
- Different users can have different style sheets: large print and text readers for example.
- The document code is reduced in size and complexity, since it does not need to contain any presentational markup.

3.4.1 User Style Sheets

Here, what user can do is they can define their own style sheets to format pages based on their preferences. For instances, people with visual impairments may want to increase the page's text size. Coding below contain an author style. The font-size is set to 9pt for all <p> tags that have class **note** applied to them.

```
<head>
        <title>User Style Sheet</title>
        <style type = "text/css">
        .note {font-size : 9pt}
        </style>
</head>
<body>
        <p> Thanks for visiting my Web site. I hope you enjoy it.
        </p><p class = "note">Please Note: This site will be moving soon. Please check periodically for updates.</p>
</body>
```

Figure 3.5: The Author Style Coding

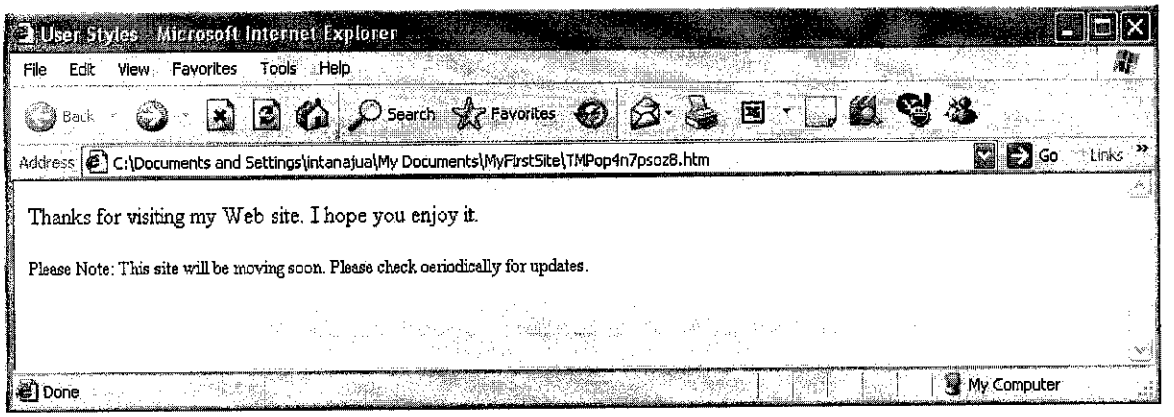


Figure 3.6: The Web Page Display for Author Style

Let says, users defined their own font-size in a User Style Sheet (20pt), color to yellow and background-color to #000080

```
/*Figure 3.5; userstyle.css*/
/* A user stylesheet*/
body    {font-size:20;
          color:yellow;
           background-color: #000080}
```

Figure 3.7: The User Style Coding

Figure 3.8: The Web Page Display for User Style

From the Figure 3.8, this is the Web page displayed when the user style sheet from Figure 3.7 is implemented in coding as shown in Figure 3.5. The User Style Sheets are external style sheet. Using this User Style Sheet, the user style is not overriding the author's defined styles. Here, it means that the User Style Sheets are not linked to documents; they are set in the browser's options.

To add a user style sheet in Internet Explorer 6, select **Internet Options** located in the **Tools** menu. In the **Internet Options** (Figure 3.9) dialog that appears, click **Accessibility**. Then,

check the **Format documents using my style sheet** checkbox, and type the location of the user style sheet. Internet Explorer 6 applies the user style sheet to any documents it loads.

3.5 Tools Required

Various tools will be used in order to assists the development project. The major tools that are going to be used are as follows:

- Macromedia Dreamweaver – to enable easier design of HTML codes and support for web development that is more comprehensive compare to other product.
- PHP and MySQL – languages to be used in order to generate the website design for the content and database design
- Adobe Photoshop 7.0 – tool used to help editing and designing graphic that meets the target audience requirements.
- Internet access – provide by the university in order to assist in the research conducted and gathering of information for the website content.

CHAPTER 4
RESULT AND DISCUSSION

4.1 Evaluation

Evaluation of the complete product is conducted among UTP students in their own residential college. The kind of evaluation is similar to field studies evaluation where this style takes the designer or evaluator out into the user's work environment in order to observe the system in action. This evaluation style allows evaluator to study the interaction and seeing the user in his 'natural environment' which is in their work or study environment. In this environment user will be exposed with interruptions and feeling as in real situation where they usually surf the Internet.

This evaluation is done in order to:

- To asses the extend of the system's functionality
- To asses the effect of the interface on the user
- To identify any specific problems with the system
- To get user's respond towards website personalization

4.2 Evaluation Procedure

The steps for conducting the testing are:

1. 20 users in taken to be as subject for the testing – consists of 2 profiles, Profile A and B (10 for each profile).
2. Each of the users is given opportunity to browse the website in the 'natural environment' that is the work/study environment.
3. After they browse the website, the users are interviewed (informal interview).

4. Each of them is given a set questionnaire to answer.

10 respondents are taken for the test, which consist of two profiles, Profile A and Profile B. 5 respondents will represent each profile. Profile A consists of Malay male students and age between 15-24 years old. Meanwhile, Profile B consists of Chinese male student age between 15-24 years old. From these two profiles we can see that the difference is in terms of their race. From this, we can make comparison on of the result from both profiles.

First, all the respondents are given opportunity to browse the website. No specific time is given for each respondent. This is done at the student's residential house, which is their natural environment as most of their time they spend in surfing the Internet are in this environment. This will give the respondents feeling of comfort and familiarity as they are used to the environment.

After they were given time to browse the website, they will be interview by the evaluator. The interview is conducted informally to the users. The objective of the interview is to get their feedback on their overall opinion regarding the system functionality and effectiveness as well as to know their satisfaction towards the website. Questions regarding their opinion on personalization technology and future enhancement for the website were also asked. The feedback get were more subjective compare to the questionnaires.

Meanwhile, after the user used the system and had undergone the interview, they were given a set of questionnaires to answer. From the questionnaires, the response gather from user were more specific and easy to analyze. The questionnaires were design in the three main sections. Sections 1 focused on user's satisfaction based on the website design. The users were asked to indicate their satisfaction in terms of the design factors such as text color and style, background color and layout. The second section focused on website personalization advantages. User asked to indicate their level of agreement on the statement provided. Meanwhile, Section 3 intention was to

get user's feedback and suggestions in the design area as in Section 1. This is important for future enhancement of the project. For detail questionnaires, please refer to the Appendices.

4.3 Result and Findings

From the evaluation conducted, all the feedback and data collected were analyzed to get the result. Following are the details of the result.

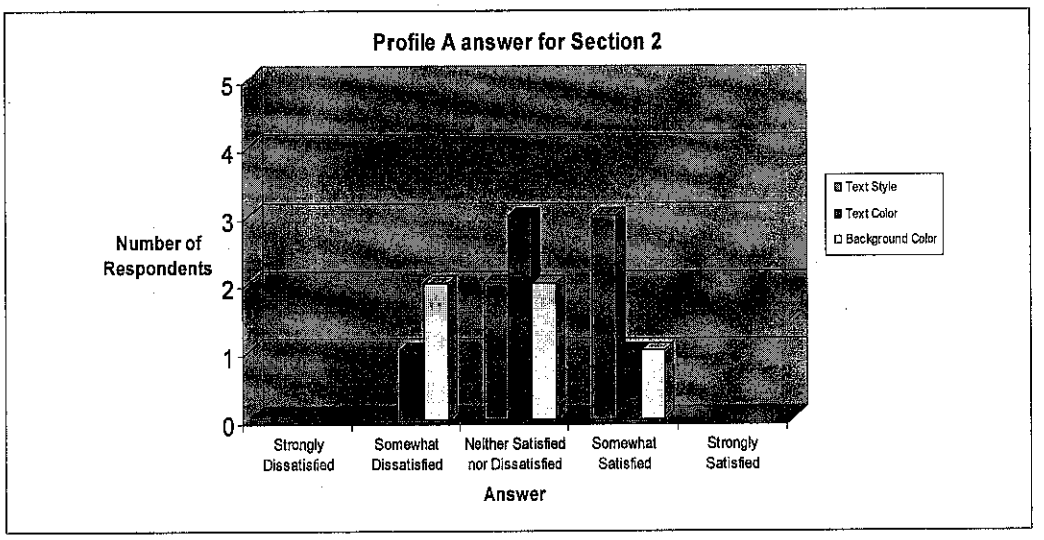


Figure 4.1: Profile A's answer in Section 2 for Text Style, Text Color and Background Color

Figure 4.1 shows the result from Profile A regarding three criteria of the website design tested, which are text style, text color and background color. Here we can see that for each of the criteria, the highest answers chose by the respondents is 'neither satisfied nor dissatisfied' and there are also respondents who are 'somewhat satisfied' with the design. For text style, 60% out of the total respondents from the profile are 'somewhat satisfied' and 40 % are in the middle which is 'neither satisfied nor dissatisfied' with the text style chose from the website. As for the text color, the result shows that 20% of the respondent from the profile are satisfied and 20% are dissatisfied with the color chose for the text in the website. 60% of the respondents

are 'neither satisfied nor dissatisfied' with the text color. Meanwhile, for the background color, only 20% of the respondent from this profile are satisfied and 40% are satisfied with the background color chose for the website design. There are also 40% of the respondents who answer neither satisfied nor dissatisfied with the text color for the background color chosen. From the result, it can be concluded that most of the respondents from profile A are neither satisfied nor dissatisfied with the text style text color and background color chose for the website.

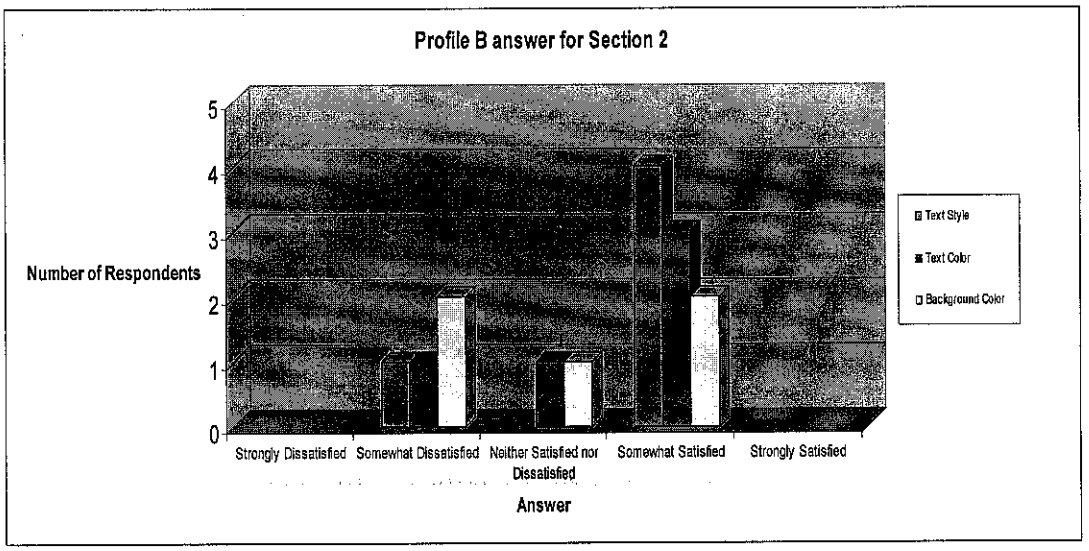


Figure 4.2: Profile B's answer in Section 2 for Text Style, Text Color and Background Color

Figure 4.2 shows the result from the profile B regarding the same three criteria of the website design tested, which are text style, text color and background color. Here we can see that respondent's responds are different with the previous result on the three criteria from Profile A. For each of the criteria, the highest answers chose by the respondents somewhat satisfied with the design. There are 20% of the respondents neither satisfied nor dissatisfied with the text color and background color from the website layout. Meanwhile, there 40% of the respondents are dissatisfied with the background color and 20% of the respondents are dissatisfied with the text style and text color of the website. Conclusion that can be made from the result, there are many

users from profile B who are satisfied with the text style, text color and background color of the website.

4.4 Discussion

Web and Internet technology is normally related to elements in Human Computer Interaction (HCI). In designing and developing a website or a system, elements in the HCI must be taken into consideration. Personalization technology is not a new technology. It has been implemented long before today. Two important element of HCI related to personalization technology will be discussed in this section.

4.4.1 User Satisfaction on the Design

User satisfaction on the design can be measures based on the evaluation result. The result is categorized in terms of text, color, layout and content.

a) Text

From the result and findings, it is concluded that most of the respondents from Profile A are neither satisfied nor dissatisfied with the text style, text color and background color of the website. On the other hand, most of the respondents from Profile B are satisfied with the text style and color chose for the website. This means that the design meet their preferences. From Figure 4.2, it is shown that 80% of the respondents is satisfied with the text style, 60% of them satisfied with the text color and 40% of the respondent are satisfied with the background color. Profile A consists of Malay male student between 15-24 years old. Meanwhile, Profile B consists of Chinese male student between 15-24 years old as well. For Profile A, the text style chose was Times New Roman, meanwhile for Profile B; the text style chose was Arial. From the two graphs above, respondent from Profile A is not very satisfied with the text style, text color and background color of the website. This is could be for the reason they want different and new style in website. As a student, respondents from Profile A and B always do their assignments and project using the default font

which is Times New Roman. Maybe from a personalized website, they are looking for another text style that is different from norm. Meanwhile, users in Profile B prefer text style Verdana which give them less formal rather than Times New Roman, and for text color, Figure 4.2 shows that 60% of the respondents from Profile B are satisfied with the color chose. This can be considered good enough since the responses from users in Profile A is not very satisfied with the text color. Meanwhile, users from Profile B prefer to have less formal colored text but still the usual color that they are familiar with, which is blue colored text.

b) Color

For color, two main focuses are the background color and combination of color for text and background in the website. Figure 4.1 shows that 60% of the users from Profile A are neither satisfied nor dissatisfied with the text color. Meanwhile 40% are somewhat satisfied with the background color. In Figure 4.2 shows that 60% of the respondents from Profile B are somewhat satisfied with the text color chose. For both Profile A and B, background color chose was white. The result shows the background color chose suit the user's preferences for Profile B only.

For color combination, Figure 4.1 shows that 60% of the total users from Profile A are neither satisfied nor dissatisfied and 20% are satisfied with the color combination. There also 20% users who are not very satisfied with the color combination. Users in Profile A preference to have dark colored text on white or light color background has changed. Meanwhile, users in Profile B prefer to have black colored text on white or light color background. These results reflect their preference for text color and background color. For Profile A, they prefer to have black color text and blue colored background. This answers suit their answers for color combination of text color and background color. The website design for them will have black colored text and light blue background color. Meanwhile for Profile B, they prefer to have black colored text and blue or other light colored background. So, the website design for them will be black colored text on light colored background.

4.4.2 Effectiveness of Personalization

On this project, effectiveness will be value highly by implementing a functional web designs that focus on to each of the users. According to the survey done earlier during the evaluation, on average, 40% to 60% of the respondents from profile A are neither agree nor disagree and 40% to 80% of the respondents from Profile B are strongly agreed that website personalization can give the users advantages such as:

- Enhance the user's surfing experience
- Enhance the interaction between users and computers
- Increase the user's interest to visit certain website
- Save user's time as user do not have to waste their time searching for certain information.

From the evaluation, 60% of the users from Profile A are somewhat agree that on overall the website suit their preferences. Meanwhile, most of users from Profile B are somewhat agree and 20% are also agreed that personalization technology does help them in providing the opportunity to tailor a site to the user's specific needs and filter out a lot of content that is not relevant to them. Some respondent also mentioned that this technology can give a lot of other advantages if properly implemented but if it is not properly implemented, it will give user's a lot of unwanted impact such as violation of privacy and too many assumptions made on one's like and dislikes.

On overall, it can be concluded that most of the users from these two profiles agree that the website using personalization technology is effective as it can give a lot of advantages to the users.

CHAPTER 5
CONCLUSION

5.1 Challenges

Upon completing the project, numbers of challenges are encountered. These challenges, internal and external, affect the research and website development process.

5.1.1 Searching Relevant Information

The duration for completion of this project is approximately about 4 months. In gathering the preliminary data for the research and website development, challenges that have been encountered are in terms of finding the relevant and reliable materials. Sources from website cannot be fully reliable except for journal and research paper.

However, due to problems in Internet connections in the campus, process of searching and downloading the materials has consumed more time than it supposed to. Besides, reference books regarding the topic are limited and problems such as books cannot be borrowed from the library can also contribute to the challenges. However, gaining and categorizing only the relevant information are very important. This step is done thoroughly to make sure of its accuracy.

5.1.2 Product Development

The development of the website is the most challenging part. The challenges lie in integrating the survey results into one product and making the product satisfies the user's preferences. Besides, the development of the product also required knowledge

of some new language such as PHP. The development process consumed more time, as the author is not familiar with the language.

5.2 Product Limitations

Personalization plays a big role in today web application such as E-Commerce, E-Business and E-Telemarketing. Besides its advantages that result of this technology either to company or individual, the whole idea is much more related to HCI elements. With regards to HCI and usage of personalization in the real world, this product has several limitations:

- Website's design that are that are build limited to Profile A and Profile B only.
- Design of the website depends solely on the web developer's prediction besides the survey data
- There is limited number of sampling and the sampling area is targeted which may create result biases.
- Evaluation done only for two profiles.

5.3 Conclusion

On this project, 60% of the users from Profile A are agreeing that on overall the website suit their preferences. Meanwhile, most of users from Profile B are agree and 20% are also agreed that personalization technology does help them in providing the opportunity to tailor a site to the user's specific needs and filter out a lot of content that is not relevant to them. The effectiveness is highly valued in implementing a functional web designs that focus on to each of the users.

In designing the website, some HCI issue is taken into consideration. Focus is given to two principles of usability, effectiveness satisfaction. These two factors are then use as a guideline in testing and evaluating the website. Evaluation of completed product was conducted with participation from users as the evaluators. The evaluation is done in the study or work environment to give them a natural and familiar environment.

As the project reached towards the end, project objectives have been achieved. User's satisfaction and effectiveness of the product cased on personalization is evaluated during the evaluation. From the result, it is concluded that by using personalization, user can have satisfaction in surfing the website and personalization technology is effectives as it can give many advantages to the users.

5.4 Recommendations for Future Enhancement

From the evaluation, respondents have suggested positive comments. Some of these comments are also from the developer's observation. The recommendations for future enhancement for the website can be concluded as below:

- Number the sampling should be increase.
- Develop all profiles graphical user interface.
- Targeted area for survey should be widens.
- Use systematic data collection methods for gathering survey result.
- Enhance elements in the website such as picture and animation.
- Increase the security level for the application.

REFERENCES

BOOK

H.M.Deitel, P.J.Deitel and A.B.Goldberg 2004. *Internet and World Wide Web: How to Program*, 3rd Edition, Pearson, Prentice Hall.

JOURNALS

[1] Massimiliano Albanese, Antonio Picariello, Carlo and Lucio Sansone, 2004, *A Web Personalization System based on Web Usage Mining Techniques,* Department of Information System, University of Napoli Federico II, Italy.

[2] Magdalini Eirinaki and Michalis Vargiannis, 2003, *Web Mining Personalization,* Athens University of Economics and Business.

[3] Harris Kravats, 2000, *Designing Web Personalization Features*

[4] Kathy Dai and Bamshad Mobasher, 2002, *A Road Map to More Effective Web Personalization: Integrating Domain Knowledge with Web Usage Mining,* School of Computer Science, Telecommunication and Information Systems, DePaul University, Chicago.

[5] Ahu Sieg, Bamshad Mobasher and Robin Burke, 2003, *Inferring User's Information Context: Integrating Use Profiles and Concept Hierarchies,* School of Computer Science, Telecommunication and Information Systems, DePaul University, Chicago.

[6] Roza Meo, Pier Luca Lanzi and Maristella Matera, 2003, *Integrating Web Conceptual Modeling and Web Usage Mining,* University of Torino, Italy.

[7] Bamshad Mobasher, 2004, *Web Usage Mining and Personalization,* School of Computer Science, Telecommunication and Information Systems, DePaul University.

[8] Bamshad Mobasher, Robert Cooley and Jaideep Srivastava, 1999, *Automatic Personalization Based on Web Usage Mining.*

[9] Kathy Dai and Bamshad Mobasher, 2003, *Integrating Semantic Knowledge with Web Usage Mining for Personalization* , School of Computer Science, Telecommunication and Information Systems, DePaul University, Chicago.

[10] Siti Sarah Harun, 2004, *Website Personalization Based-On Demographic Data,* Degree in Information Technology Final Year Project Report, University Technology of PETRONAS, Perak,Malaysia.

APPENDICES

1.0 Survey Questionnaires

SURVEY: WEBSITE PERSONALIZATION BASED ON DEMOGRAPHIC DATA

SECTION 1: PERSONAL DETAILS

1. What is your age range?
   A. 15 – 24 B. 25 – 34 C. 35 – 44 D. 45 and above

2. What is your gender?
   A. Male B. Female

3. What is your race?
   A. Malay B. Chinese C. Indian D. Others

4. Which title best describes your job description?
   A. Executive B. Manager C. Technical D. Student

SECTION 2: INTERNET USAGE AND KNOWLEDGE

1. How many days a week, on average, do you surf the web (not including the E-mail and Friendster)?
   A. 7- 6 days B. 5 – 4 days C. 3 – 2 days D. 1 – 0 day

2. How many hours a day, on average, do you surf the web (not including E-mail and Friendster)?
   A. 0–2 hours B. 3–5 hours C. 6-8 hours D. 9++ hours

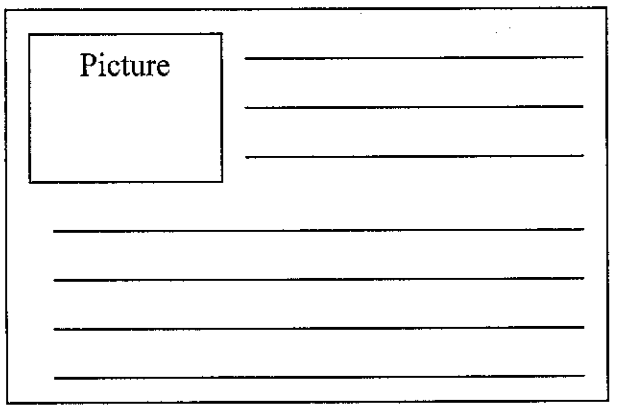
3. Which source(s) do you usually use when searching for information?
   A. Internet B. Library C. Both D. Other:_____

4. Are you familiar with online health services (e.g. telemedicine consultant)
   A. Yes B. No

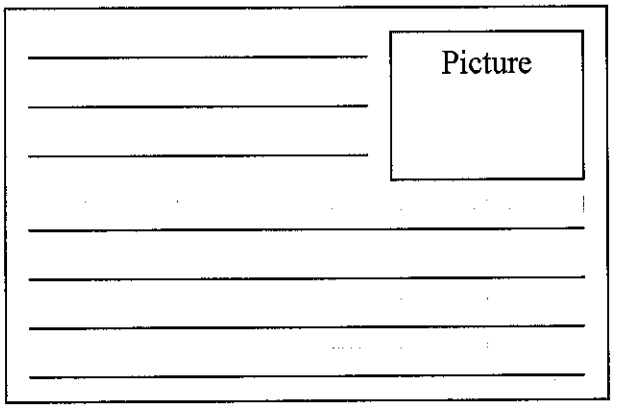
SECTION 3: PICTURE

Which picture position do you prefer most in a website?

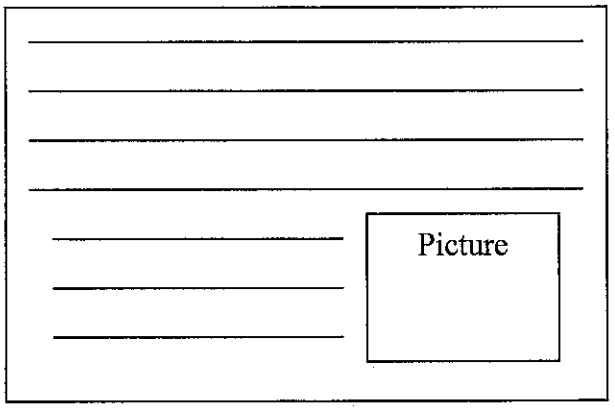
A.



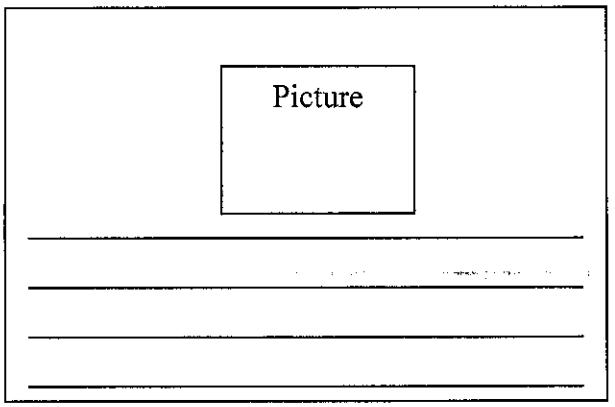
B.



C.

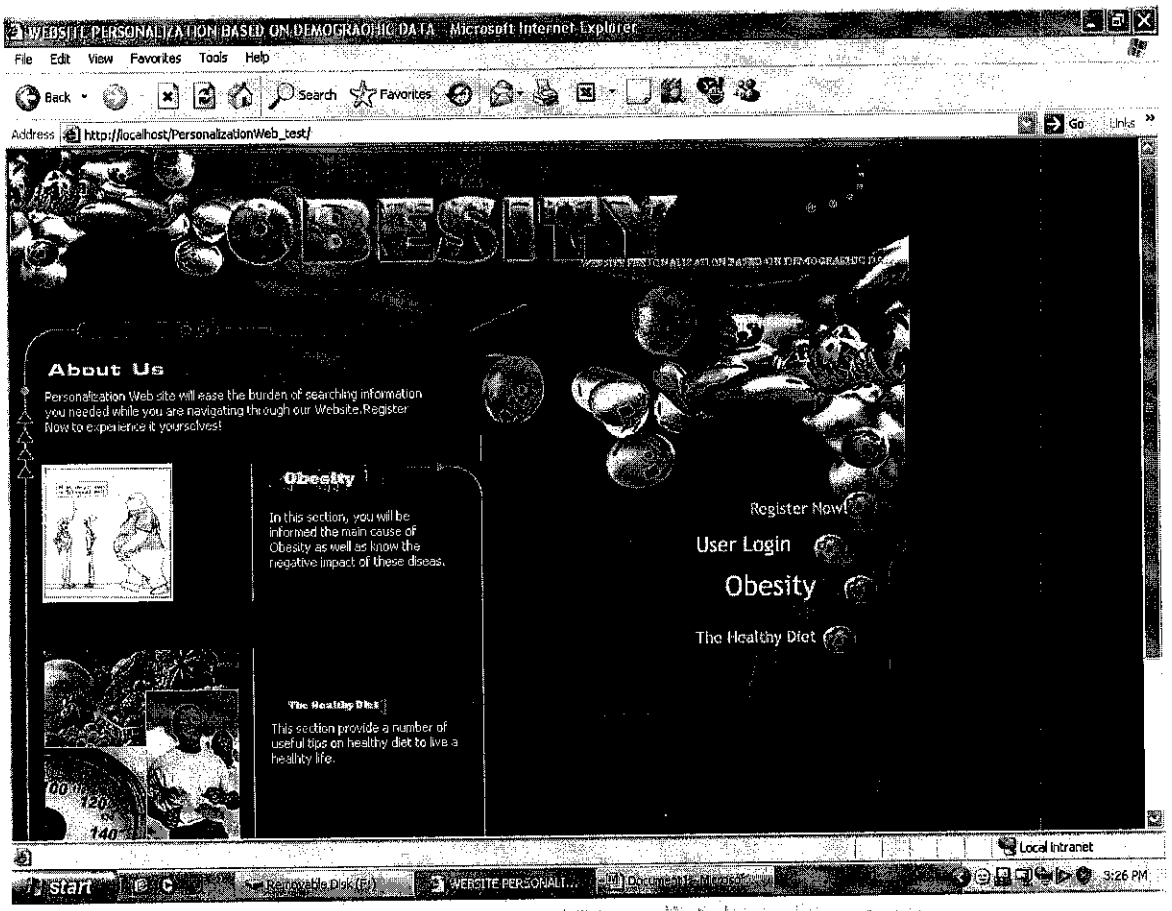


D.

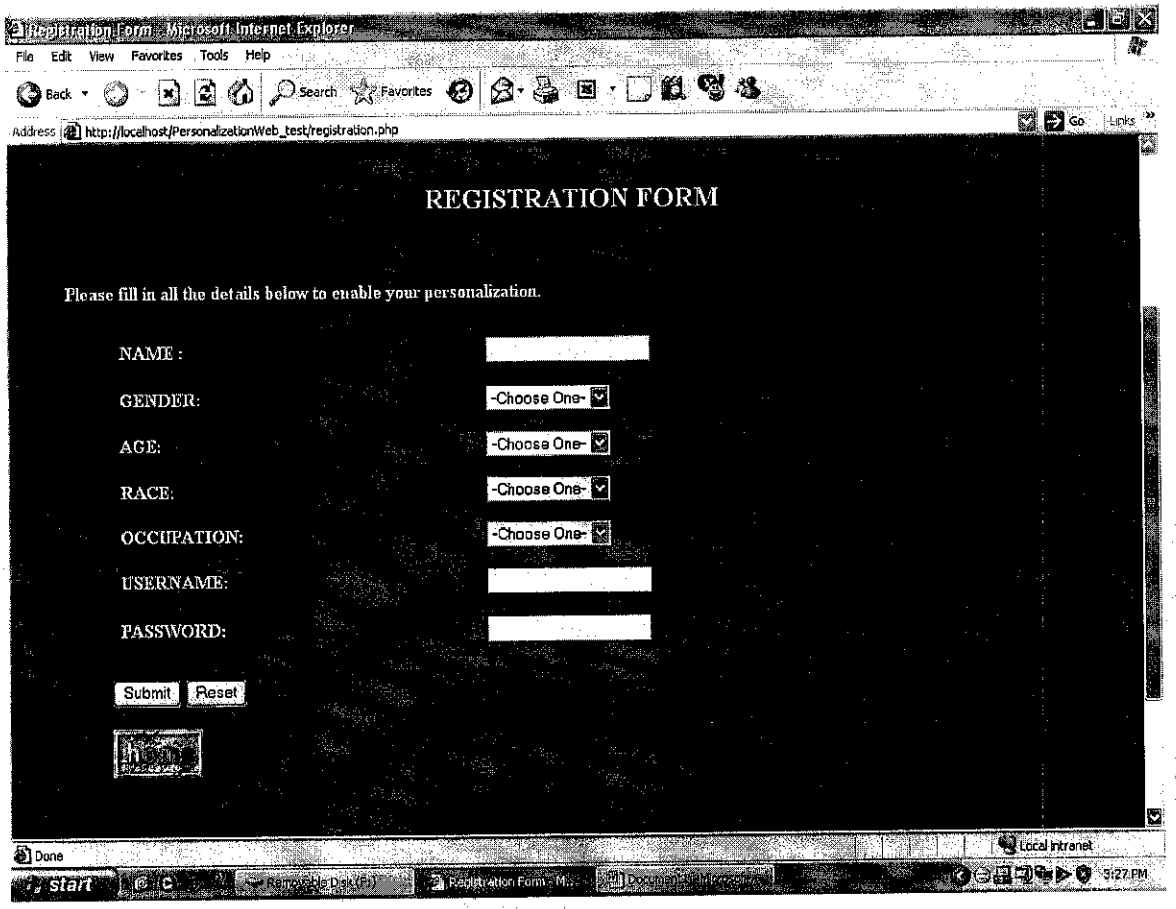


2.0 User Interface Design

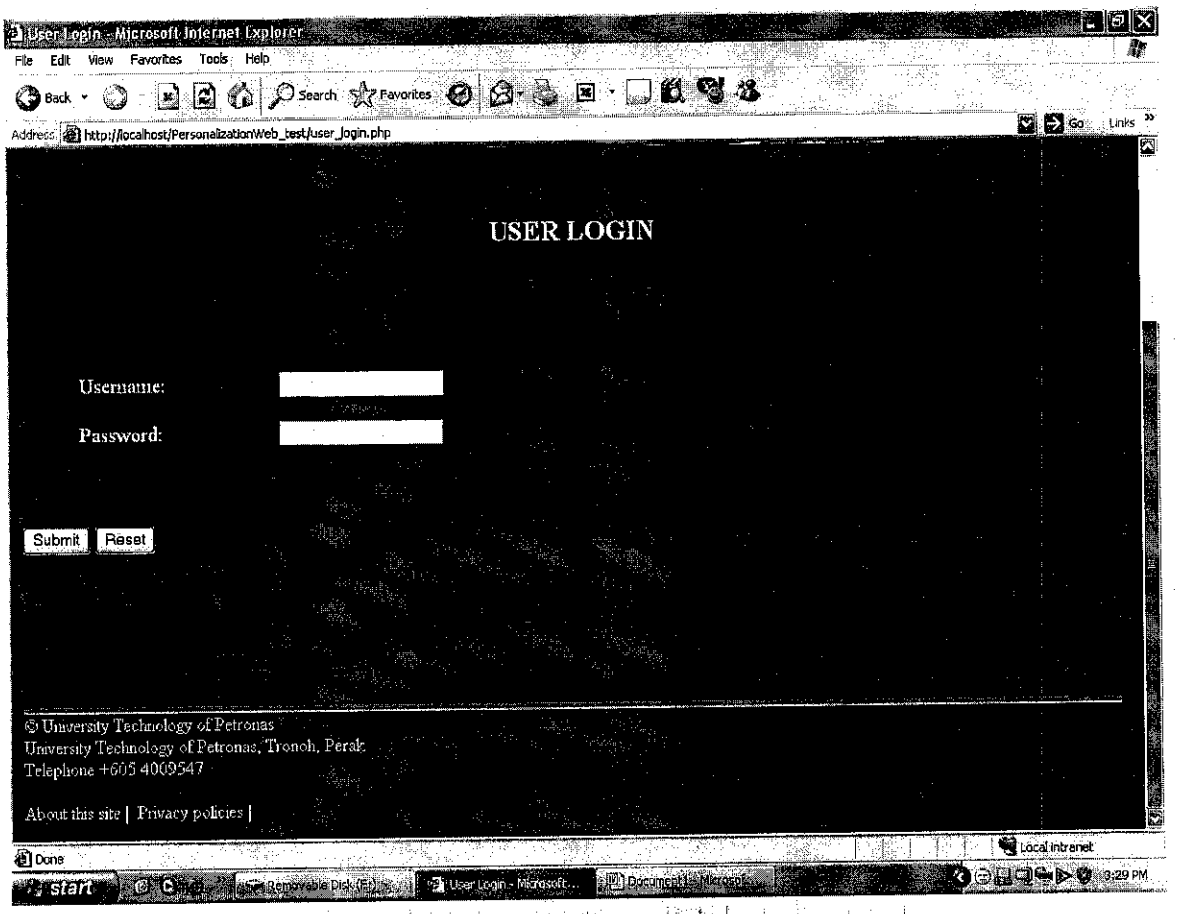
2.1 Index Page



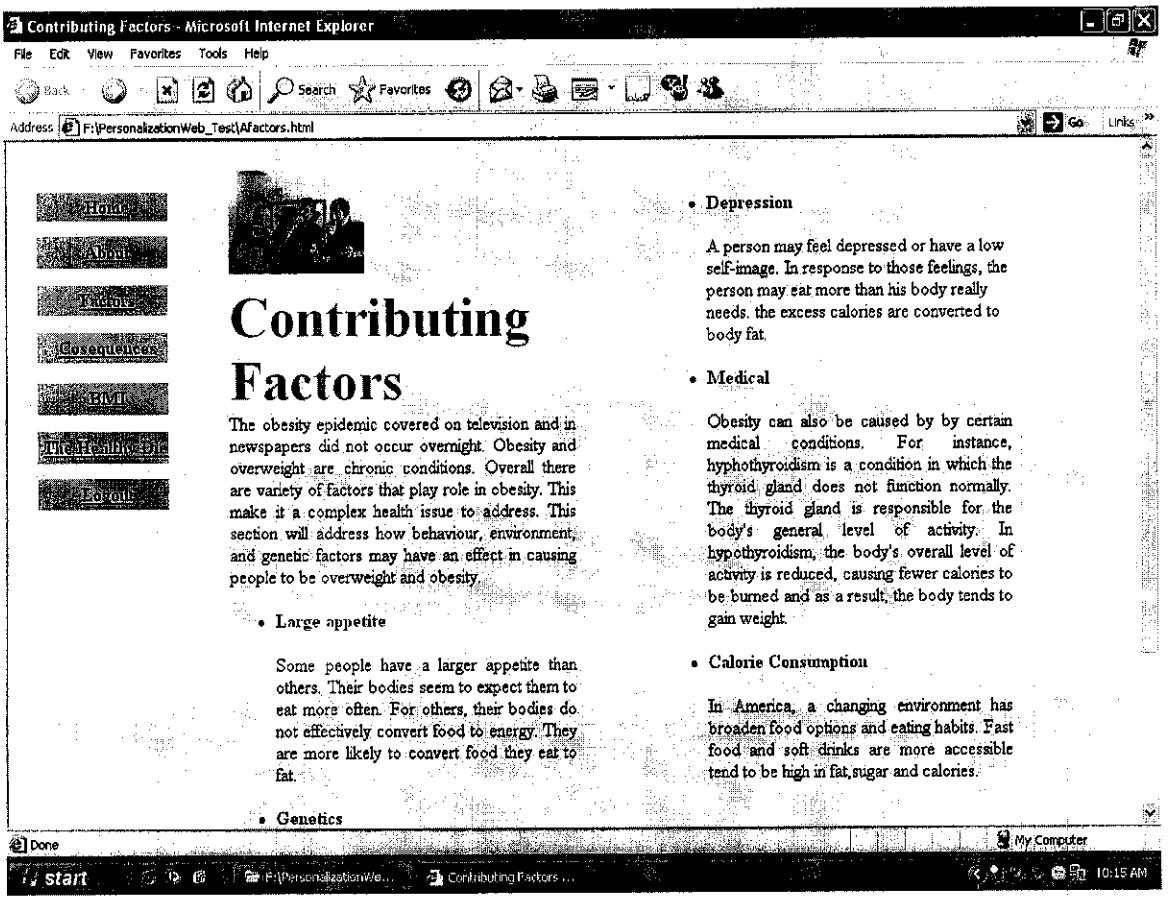
2.2 Registration Page



2.3 Login Page



2.4 Profile A's User Interface Design



2.5 Profile B's User Interface Design