# The Influence of Human Factors on Users' Preferences of Web-Based Applications: A Data Mining Approach

A Thesis submitted for the degree of Doctor of Philosophy by

Natalie Christine Clewley

School of Information Systems, Computing and Mathematics

Brunel University

September 2010

# Abstract

As the Web is fast becoming an integral feature in many of our daily lives, designers are faced with the challenge of designing Web-based applications for an increasingly diverse user group. In order to develop applications that successfully meet the needs of this user group, designers have to understand the influence of human factors upon users' needs and preferences. To address this issue, this thesis presents an investigation that analyses the influence of three human factors, including cognitive style, prior knowledge and gender differences, on users' preferences for Web-based applications. In particular, two applications are studied: Web search tools and Web-based instruction tools.

Previous research has suggested a number of relationships between these three human factors, so this thesis was driven by three research questions. Firstly, to what extent is the similarity between the two cognitive style dimensions of Witkin's Field Dependence/Independence and Pask's Holism/Serialism? Secondly, to what extent do computer experts have the same preferences as Internet experts and computer novices have the same preferences as Internet novices? Finally, to what extent are Field Independent users, experts and males alike, and Field Dependent users, novices and females alike? As traditional statistical analysis methods would struggle to effectively capture such relationships, this thesis proposes an integrated data mining approach that combines feature selection and decision trees to effectively capture users' preferences. From this, a framework is developed that integrates the combined effect of the three human factors and can be used to inform system designers.

The findings suggest that firstly, there are links between these three human factors. In terms of cognitive style, the relationship between Field Dependent users and Holists can be seen more clearly than the relationship between Field Independent users and Serialists. In terms of prior knowledge, although it is shown that there is a link between computer experience and Internet experience, computer experts are shown to have similar preferences to Internet novices. In terms of the relationship between all three human factors, the results of this study highlighted that the links between cognitive style and gender and between cognitive style and system experience were found to be stronger than the relationship between system experience and gender.

This work contributes both theory and methodology to multiple academic communities, including human-computer interaction, information retrieval and data mining. In terms of theory, it has helped to deepen the understanding of the effects of single and multiple human factors on users' preferences for Web-based applications. In terms of methodology, an integrated data mining analysis approach was proposed and was shown that is able to capture users' preferences.

# Publications

The following publications have resulted from work conducted related to the investigation undertaken in this thesis.

**Journal Papers**

- Clewley, N., Chen, S. Y., and Liu, X. (2010). Cognitive styles and search engine preferences: Field dependence/independence vs. holism/serialism. *Journal of Documentation*, Vol. 66 No.: 4, pp. 585-603.
- Clewley, N., Chen, S. Y., and Liu, X. (Accepted). Mining the Relationship between Cognitive Styles and Web-based Instruction: Holists vs. Serialists. *Educational Technology and Society*.
- Clewley, N., Chen, S. Y., and Liu, X. (2009) Evaluation of the credibility of internet shopping in the UK, *Online Information Review*, Vol. 33 No. 4, pp. 805-826.

**Conference Papers**

- Clewley, N., Chen, S. Y., and Liu, X. (2009). Cognitive styles and web-based instruction: field dependent/independent vs. Holist/Serialist. In *Proceedings of the 2009 IEEE international Conference on Systems, Man and Cybernetics*, San Antonio, TX, USA, October 11 - 14, 2009.

**Book Chapters**

- Clewley, N., Chen, S. Y., and Liu, X. (2009). Applications for Data Mining Techniques in Customer Relationship Management, *Encyclopedia of Information Science and Technology*, 2nd Edition. Idea Group Publishing,

# Acknowledgements

There are a number of people for whom I am very grateful and without their support this thesis would not have been possible. My first thanks goes to both my supervisors, Professors Sherry Y. Chen and Xiaohui Liu. Their continued support, encouragement and guidance has challenged my thinking and allowed me to develop both as a person and as a researcher.

I am forever grateful to the School of Information Systems, Computing and Mathematics for giving me the opportunity to study for this PhD and also to my fellow members of the Centre of Intelligent Data Analysis group for making the office a warm and friendly environment.

# Table of Contents

**Abstract** ..... **i**
**Publications** ..... **ii**
**Acknowledgements** ..... **iii**
**Table of Contents** ..... **iv**
**List of Figures** ..... **viii**
**List of Tables** ..... **ix**
**List of Abbreviations and Acronyms** ..... **xi**

**Chapter 1 – Introduction** ..... **1**
1.1 Context ..... 1
1.2 Motivations ..... 2
1.3 Aims and Research Questions ..... 3
1.4 Contributions ..... 4
1.5 Structure of Thesis ..... 5

**Chapter 2 – Related Literature** ..... **6**
2.1. Introduction ..... 6
2.2. Web-based Applications ..... 7
2.2.1 Web Search Tools ..... 7
2.2.2 Web-Based Instruction Tools ..... 7
2.3. Human Factors ..... 8
2.3.1. Cognitive Style ..... 8
2.3.1.1 Field Dependence/Independence ..... 9
2.3.1.2 Holist/Serialist ..... 13
2.3.2 Prior Knowledge ..... 15
2.3.3 Gender Differences ..... 18
2.3.4 Cognitive Style vs. Prior Knowledge vs. Gender Differences ..... 21
2.4 Analysis of Human Factors ..... 22
2.4.1 Classification ..... 24
2.4.2 Clustering ..... 25
2.4.3 Association Rules ..... 25
2.4.4 Information Visualisation ..... 26
2.4.5 Feature Selection ..... 27
2.5 Summary ..... 28

**Chapter 3 – Study 1: Field Dependent/Independent vs. Holist/Serialist............30**
3.1 Introduction ........................................................................................30
3.2 Data Collection: Dataset 1 – Web Search Tools .......................................31
3.2.1 Participants ........................................................................................31
3.2.2 Research Instruments ..........................................................................32
3.2.2.1 Cognitive Styles Analysis (CSA) ......................................................32
3.2.2.2 Study Preference Questionnaire (SPQ) ............................................33
3.2.2.3 Exit Questionnaire .........................................................................34
3.2.3 Experimental Procedure ......................................................................35
3.3 Data Collection: Dataset 2 – Web-Based Instructional Tools .....................36
3.3.1. Participants ........................................................................................36
3.3.2. Research Instruments ..........................................................................37
3.3.3. Experimental Procedure ......................................................................38
3.4 Data Analysis ..........................................................................................39
3.4.1 Feature Selection ................................................................................40
3.4.2 Decision Tree Classification ................................................................42
3.5 Findings: Dataset 1 ..................................................................................42
3.5.1. Feature Selection ................................................................................42
3.5.1.1. Field Dependence/Independence ....................................................43
3.5.1.2. Holist/Serialist ..............................................................................43
3.5.2. Decision Tree Classification ................................................................44
3.5.2.1 Field Dependent/Independent ..........................................................46
3.5.2.2 Holist/Serialist ...............................................................................48
3.5.2.3 Field Dependent/Independent vs. Holist/Serialist ............................51
3.5.3 Statistical Significance ........................................................................52
3.5.4 Summary ............................................................................................53
3.6 Findings: Dataset 2 – Web-Based Instruction Tool ....................................54
3.6.2 Feature Selection ................................................................................54
3.6.2.1 Field Dependent/Independent ..........................................................54
3.6.2.2 Holist/Serialist ...............................................................................54
3.6.2.3 Field Dependent/Independent vs. Holist/Serialist ............................55
3.6.3 Decision Tree Classification ................................................................55
3.6.4 Statistical Significance ........................................................................60
3.6.5 Summary ............................................................................................61
3.7 Conclusion ..............................................................................................62

**Chapter 4 – Study 2: Internet Experience vs. Computer Experience ................. 63**
4.1. Introduction ........................................................................................ 63
4.2. Data Collection ...................................................................................... 64
4.2.1 Personal Information Sheet ................................................................ 64
4.2.1.1 Dataset 1: Web Search Tools ........................................................ 65
4.2.1.2 Dataset 2: Web-Based Instruction Tools ........................................ 65
4.3 Data Analysis ........................................................................................ 66
4.4 Findings – Dataset 1: Web Search Tools ................................................. 66
4.4.1 Feature Selection ................................................................................ 68
4.4.1.1 Internet Experience ......................................................................... 68
4.4.1.2 Computer Experience ...................................................................... 68
4.4.1.3 Internet Experience vs. Computer Experience ................................ 69
4.4.2 Decision Tree Classification ............................................................... 69
4.4.2.1 Internet Experience ......................................................................... 69
4.4.2.2 Computer Experience ...................................................................... 72
4.4.2.3 Internet Experience vs. Computer Experience ................................ 73
4.4.3 Statistical Analysis ............................................................................. 76
4.5 Findings – Dataset 2: Web-Based Instruction Tools ................................. 77
4.5.1 Feature Selection ................................................................................ 77
4.5.1.1 Internet Experience ......................................................................... 77
4.5.1.2 Computer Experience ...................................................................... 77
4.5.1.3 Internet Experience vs. Computer Experience ................................ 78
4.5.2 Decision Tree Classification ............................................................... 79
4.5.2.1 Internet Experience ......................................................................... 79
4.5.2.2 Computer Experience ...................................................................... 81
4.5.2.3 Internet Experience vs. Computer Experience ................................ 83
4.5.3 Statistical Significance ........................................................................ 84
4.6 Findings: Dataset 1 vs. Dataset 2 ............................................................ 85
4.6.1 Internet Experience ............................................................................. 85
4.6.2 Computer Experience .......................................................................... 85
4.6.3 Internet vs. Computer Experience ....................................................... 85
4.7 Summary ................................................................................................ 86

**Chapter 5: Study 3 – Cognitive Style vs. Previous Experience vs. Gender ........ 88**
5.1 Introduction ........................................................................................... 88
5.2 Data Collection ...................................................................................... 90
5.2.1 Personal Information Sheet ................................................................ 90

5.3 Results: Dataset 1 – Web Search Tools ... 90
5.3.1 Feature Selection ... 90
5.3.2 Decision Tree Classification ... 91
5.3.3 Statistical Significance ... 94
5.4 Results: Dataset 2 – Web-Based Instruction Tools ... 94
5.4.1 Feature Selection ... 94
5.4.2 Decision Tree Classification ... 95
5.4.3 Statistical Significance ... 97
5.5 Dataset 1 vs. Dataset 2 ... 97
5.6 Results: Cognitive Style vs. Previous Experience vs. Gender ... 98
5.6.1 Dataset 1 – Web Searching ... 98
5.6.2 Dataset 2 – Web-Based Instruction ... 102
5.6.3 Dataset 1 vs. Dataset 2 ... 104

**Chapter 6 – Development of an Integrated Model ... 106**
6.1. Introduction ... 106
6.2. Development of an Integrated Model ... 106
6.2.1 Dataset 1: Web Search Tools ... 106
6.2.2 Dataset 2: Web-based Instruction Tools ... 108
6.3. Implications for System Design ... 113
6.4. Benefactors ... 113

**Chapter 7 – Conclusion ... 117**
7.1 Introduction ... 117
7.2 Investigating the Relationships Between Human Factors ... 117
7.2.1 Research Question 1 ... 117
7.2.2 Research Question 2 ... 118
7.2.3 Research Question 3 ... 119
7.3 Identifying a Suitable Analysis Method ... 119
7.4 Significance of Thesis ... 120
7.5 Limitations and Suggestions for Future Work ... 121

**References ... 123**

**Appendices ... 136**
Appendix A ... 137
Appendix B ... 141

# List of Figures

Figure 3.1: The framework of Study 1 ....31
Figure 3.2. Dataset 1 - Participants' Cognitive Style Distributions ....34
Figure 3.3. Dataset 2 - Participants' Cognitive Style Distributions ....37
Figure 3.4: WBI Interface....39
Figure 3.5. Data Analysis Framework for Study 1 ....40
Figure 3.6. Four Step Feature Selection Method....41
Figure 3.7: Dataset 1 - Field Dependent/Independent Decision Tree....47
Figure 3.8: Dataset 1 Holist/Serialist Decision Tree ....49
Figure 3.9. Dataset 2 - Field Dependent/Independent Decision Tree....57
Figure 3.10. Dataset 2 - Holist/Serialist Decision Tree 1 ....59
Figure 3.11. Dataset 2 - Holist/Serialist Decision Tree 2 ....59
Figure 3.12. Relationship between Field Dependence/Independence and Holism/Serialism.62
Figure 4.1: Framework for Study 2 ....64
Figure 4.2: Likert scale for rating Internet and computer experience....65
Figure 4.3. Data Analysis Framework for Study 2 ....67
Figure 4.4. Dataset 1 - Internet Experience Decision Tree....71
Figure 4.5. Dataset 1 – Computer Experience Decision Tree ....74
Figure 4.6: Dataset 2 - Internet Experience Decision Tree....80
Figure 4.7: Dataset 2 - Computer Experience Decision Tree ....82
Figure 4.8. Relationship between Internet Experience and Computer Experience ....87
Figure 5.1: Study 3 Framework....89
Figure 5.2. Participants' Gender Distributions....90
Figure 5.3: Gender - Dataset 1 Decision Tree ....92
Figure 5.4: Gender - Dataset 2 Decision Tree ....96
Figure 5.5: Relationship between Cognitive Style, Gender, System Experience....105
Figure 6.1. Integrated Model of the Influence of Human Factors on Users' Preferences of Web Search Tools....111
Figure 6.2. Integrated Model of the Influence of Human Factors on Users' Preferences of Web-Based Learning Tools ....112

# List of Tables

Table 2.1. Differences between FD/FI characteristics (Witkin *et al*., 1977) .......................... 10
Table 2.2. Differences between Holist and Serialist characteristics (Pask, 1979) ................. 14
Table 2.3. Differences Between Novice and Expert Characteristics ...................................... 17
Table 2.4: Differences between male and female characteristics ........................................... 20
Table 2.5. Statistical Methods of Analysis in Previous Studies ............................................ 23
Table 3.1. Examples of the questions in the questionnaire ..................................................... 35
Table 3.2: Classifier Families ................................................................................................... 41
Table 3.3: Dataset 1: Number of Features Selected per Family ............................................ 45
Table 3.4: Commonly Selected Features .................................................................................. 45
Table 3.5. Field Dependent/Independent (CSA) Classification Accuracies ........................... 46
Table 3.6. Decision Rules for Field Dependent/Independent Decision Tree ......................... 47
Table 3.7: Holist/Serialist (SPQ) Classification Accuracies ................................................. 49
Table 3.8. Decision Rules for Holist/Serialist Decision Tree ................................................. 50
Table 3.9. Distribution of Participants' Cognitive Styles ...................................................... 52
Table 3.10. Statistical Significance of Features (p<0.05) ...................................................... 53
Table 3.11: Dataset 2 - Number of Features Selected per Family ......................................... 55
Table 3.12: Field Dependent/Independent (CSA) Classification Accuracies ......................... 56
Table 3.13. Decision Rules for Field Dependent/Independent Tree ...................................... 57
Table 3.14: Holist/Serialist (SPQ) Classification Accuracies ................................................ 59
Table 3.15. Decision Rule Table for Holist/Serialist Trees .................................................... 60
Table 3.16. Distribution of Participants' Cognitive Styles .................................................... 60
Table 3.17: Statistical Significance of Dataset 2 .................................................................... 61
Table 4.1. Dataset 1: Participants' Level of Experience ......................................................... 65
Table 4.2. Dataset 2 - Participants' Level of Experience ...................................................... 66
Table 4.3. Dataset 1: Number of Features Selected per Family ............................................ 68
Table 4.4. Internet Experience Classification Accuracy (%) ................................................. 69
Table 4.5. Decision Rules for the Internet Experience Decision Tree ................................... 71
Table 4.6. Computer Experience Classification Accuracy (%) ............................................. 72
Table 4.7. Decision Rules for Computer Experience Tree ..................................................... 75
Table 4.8: Dataset 2: Number of Features Selected per Family ............................................ 78
Table 4.9. Internet Experience Classification Accuracy (%) ................................................. 79
Table 4.10. Decision Tree Rules for Internet Experience Tree ............................................. 81
Table 4.11. Computer Experience Classification Accuracy ................................................... 81
Table 4.12. Decision Tree Rules for Computer Experience Decision Tree ............................ 82

Table 5.1: Dataset 1: Number of Features Selected per Family ............................................91
Table 5.2. Decision Rules for Gender Decision Tree ............................................................93
Table 5.3. Statistical Significance of Features for Gender .....................................................94
Table 5.4: Dataset 2: Number of Features Selected per Family ............................................95
Table 5.5. Statistical Significance for Gender .........................................................................97
Table 5.6: Dataset 1 - Table of Commonly Selected Features ............................................100
Table 5.7. Dataset 1 - Classification Results ........................................................................100
Table 5.8. Dataset 2 - Table of Commonly Selected Features.............................................103
Table 6.1. Design Implications for Web-Based Applications .............................................115

# List of Abbreviations and Acronyms

| Acronym | Explanation |
|---|---|
| AODE | Averaged One Dependent Estimates |
| BN | Bayesian Network |
| BNC | Bayesian Network Classifier |
| CE | Computer Experience |
| F | Female |
| FD | Field Dependent |
| FD/FI | Field Dependent/Field Independent |
| FI | Field Independent |
| H | Holist |
| H/S | Holist/Serialist |
| HCI | Human-Computer Interaction |
| IE | Internet Experience |
| K* | *k*-Star |
| KNN | *k*-Nearest Neighbour |
| M | Male |
| NB | Naive Bayes |
| NN | Nearest Neighbour |
| S | Serialist |
| WBA | Web-Based Application |
| WBI | Web-Based Instruction |

# Chapter 1 – Introduction

This chapter will introduce the thesis context, the motivations and state the research aims of the thesis. The contributions of this work will be clearly stated and the structure of the thesis will be detailed.

## 1.1 Context

This thesis presents a work that encompasses multiple disciplines, touching upon elements from data mining, digital learning, human-computer interaction (HCI) and information retrieval. Whilst attempting to provide some answers to key questions that have eluded researchers in the field of HCI, a novel data mining analysis approach is employed that integrates feature selection and data mining techniques. The following section provides an introduction to each of these research areas.

The field of human-computer interaction is concerned with how different users interact with computer applications (Frias-Martinez, *et al.*, 2007). Its aim is to provide more user-friendly systems that are efficient and effective, with popular studies choosing Web search engines and Web-based instruction (WBI) tools. An important vein of HCI research is that of human factors, those individual human characteristics that affect our interaction with computer systems (Treu, 1994). Among many human factors, previous research suggests that cognitive style (Chen and Macredie, 2004; Ford, Miller and Moss, 2005), previous experience (Mitchell, et al., 2005; Liaw and Huang, 2006; Castañeda, Muñoz-Leiva, and Luque, 2007) and gender differences (Roy and Chi, 2003; Ford and Miller, 1996) significantly influence users' interactions with Web-based applications. Research into the relationship between these factors and their effect on users' preferences can help us to understand and develop more effective Web-based applications.

The field of data mining overlaps with areas in many other fields, including information systems, statistics and mathematical sciences (Bohen *et al*., 2003). Data mining is the search for valuable information within large volumes of data (Hand, Mannila, & Smyth, 2001). This valuable information can then be used to predict, model or identify interrelationships (Urtubia *et al*., 2007) without the need to predefine underlying relationships between dependent and independent variables (Chang and Chen, 2005). When current datasets can include hundreds or even thousands of data items, or features, this is a very useful technique. However, human preferences data is often ‘noisy’ and full of inaccurate information, which can lead to biased results. Therefore, a pre-processing step, such as feature selection, is necessary to sift through the data so that only the most relevant subsets are included in the mining process (Crysostomou, *et al*., 2008).

The aim of this chapter is to introduce the areas of investigation, beginning in Section 1.2 with defining the problem area. Following on from that, Section 1.3 outlines the main aims and research questions investigated in this thesis. Section 1.4 describes the contributions of this thesis before detailing the thesis structure.

## 1.2 Motivations

As the Web is fast becoming an integral feature in our daily lives, more and more users are using Web-based applications to support their online tasks. There are many different Web-based applications, with Web search engines and Web-based instruction (WBI) programs being two of the most studied in previous research. With the Web becoming increasingly accessible to a wider audience, a broad range of users from all over the globe are now using the same applications. Each of these users has a different background, different skills and levels of previous experience, and thus will have different expectations and needs when using Web-based applications. This diversity presents a challenge to designers as the effectiveness of such applications relies upon understanding and providing interfaces that support each user’s needs and preferences. As mentioned above, cognitive style, previous experience and gender differences have been shown to have a significant influence on users’ preferences. These human factors have been widely studied in previous research. However, the majority of these studies have concentrated on a single human factor, with the researchers rarely investigating the relationship between the three. Every user will always have all three to some extent, so it is vital that their relationship, if any, is ascertained so that we can identify the combined effect they have on users’ preferences. Some researchers have proposed theoretical links between these three human factors, both in terms of dimensions within human factors (i.e. Field Dependent = Holist; Field Independent = Serialist; Internet expert = computer expert; Internet novice = computer novice) and in terms of the relationship between human factors (i.e. Field Independent = expert = male; Field Dependent = novice = female). The significance of an investigation that aims to identify such relationships would not only help to inform designers to develop more effective applications, but we will also gain a deeper understanding of human factors and the interactions between them.

Many previous studies have used traditional statistical methods (e.g. ANOVA) to analyse the relationship between human factors and user preferences (e.g. DeTure, 2004). Human factors data is by nature very 'noisy', meaning that it can contain data items that are considered outliers or inaccurate data. The aforementioned statistical methods are not designed to cope with this issue, neither with the large and ever increasing amount of data that is often associated with human factors. Additionally, these methods require predefined assumptions to be made about dependent and independent variables, which could introduce bias if not accurate, especially if not much is known about the relationships between the variables beforehand. Furthermore, these methods are not able to cope with the analysis of multiple human factors at the same time. Therefore, a new method is needed that can overcome these issues.

## 1.3 Aims and Research Questions

The main aims of this thesis are twofold:

1) To investigate the relationship between the three human factors, cognitive style, system experience and gender differences using users' preferences of Web-based applications.

2) To identify a suitable methodology that can a) overcome the natural fuzziness of human preferences data, b) avoid the need to make predefined assumptions about the data, and c) have the ability to compare multiple human factors at the same time.

These aims will be accomplished through investigating a number of research questions. The three research questions (RQ) are as follows:

a) In terms of cognitive style, to what extent is the similarity between the two dimensions Witkin's Field Dependence/Independence and Pask's Holism/ Serialism? (RQ1)

b) In terms of system experience, to what extent do computer experts have the same preferences as Internet experts and computer novices have the same preferences as Internet novices? (RQ2)

c) In terms of the relationship between the three human factors, to what extent are Field Independent users, experts and males alike, and Field Dependent users, novices and females alike? (RQ3)

## 1.4 Contributions

Due to the multidisciplinary nature of this study, this thesis makes a contribution to two different communities, including human factors and data mining. These contributions are described below.

- **Human Factors (HCI)**

  With regards to the human factors community, this thesis presents a work that analyses not only single human factors, but also the relationships between human factors. This in turn provides a deeper understanding of each human factor and also the interactions between human factors. In this way, this work helps to provide answers to some key questions within the field and provide more detailed guidance for designers of Web-based applications.

- **Data Mining and Feature Selection**

  This thesis presents a novel method that integrates feature selection and decision tree classification. In addition to proving that this method can effectively capture user preferences, it is also able to effectively examine the relationships between multiple human factors. This work will also add to the body of understanding the nature of different classifiers and decision tree algorithms.

## 1.5 Structure of Thesis

Following on from this introductory chapter, Chapter 2 presents a detailed review of the three human factors, cognitive style, system experience and gender differences, and how they influence users' preferences for Web-based applications. In particular, two types of Web-based application will be examined: Web search engines and Web-based instruction programs (WBI). In addition, modern data mining and feature selection techniques will be introduced and their strengths and weaknesses examined.

Chapter 3 presents the first of the three studies analysed in this thesis: cognitive style. In this study, the effect of cognitive style on users' preferences will be investigated. More specifically, Witkin's Field Dependence/Independence will be compared with Pask's Holism/Serialism with the aim of discovering their similarities and differences. This chapter aims to answer RQ1.

Chapter 4 presents the second of the three studies analysed in this thesis: system experience. In this study, the effect of system experience on users' preferences will be investigated by comparing Internet and computer experience. This chapter aims to answer RQ2.

Chapter 5 presents the third and final study analysed in this thesis and combines the results from Chapter 3, cognitive style, and Chapter 4, system experience, with gender differences. This study aims to discover the relationship between the three human factors by identifying their similarities and differences. This chapter aims to answer RQ3.

After summarising the results from the three studies presented in this thesis, Chapter 6 presents a framework for designers of Web-based applications. This framework provides support for the considerations of the three human factors examined in this thesis. Such a framework is valuable in the development of effective Web-based applications.

Finally, Chapter 7 presents the conclusions for this thesis with regard to the influence of human factors on user preferences of Web-based applications. This chapter will also discuss limitations of these studies and propose directions for future research.

# Chapter 2 – Related Literature

## 2.1. Introduction

Having previously given an introduction to the aims of this thesis, this chapter will present a detailed review of past research regarding the influence of human factors on users' preferences of Web-based applications. In particular, three human factors, i.e., cognitive style, system experience and gender differences, and two Web-based applications, i.e., Web search engines and Web-based instruction (WBI) tools, will be included in this review. Furthermore, modern data mining and feature selection techniques will be introduced and their benefits and weaknesses explored.

The aims of this chapter are therefore:

- To review, analyse and present the current and existing literature regarding human factors and their effect on users' preferences of Web-based applications.

- To review the different analysis techniques used in current literature to examine the effects of human factors on user's preferences of Web-based applications.

- To review, analyse and present the current and existing literature regarding data mining and feature selection techniques.

- To discover a data mining framework that can be used to analyse the effect of human factors on user preferences for Web-based applications.

This chapter is organised as follows. Firstly, the chapter starts by defining Web-based applications and describes the challenges faced by designers in regards to users' preferences. More specifically, two types of Web-based application are introduced: Web search engines and Web-based instruction (WBI) tools. Subsequently, the relationships between Web-based applications, human factors and user preferences are then introduced and the three human factors analysed in this study are explored. The following section then describes past and current methods of analysing the effects of human factors on user preferences, before reviewing and presenting different data mining techniques that could be used in the analysis stage.

## 2.2. Web-based Applications

In the last two decades, we have seen the World Wide Web grow and evolve to the point where it has now become a primary source of information in our day-to-day lives (Bar-Ilan, 2005). Modern advancements in technologies mean that users now rely on the Web for an increasingly broad range of tasks. There have been many applications designed to make such tasks easier on the user, with two of the most widely studied being Web search tools and Web-based instruction tools. The next two sections describe their purposes and the current challenges that designers face.

### 2.2.1 Web Search Tools

The Web enables the wide dissemination of information and services (Ivory and Megraw, 2005) and has evolved to become the largest man-made repository of information with more than tens of billions of pages (Baeza-Yates and Raghavan, 2010). Since the middle of the 1990s, there has been an increasing trend toward reliance on the Web for information needs, with the Web becoming a major information source in our daily life (Bar-Ilan, 2005). In particular, millions of people use search engines every day to locate information (Spink *et al.*, 2006). Search engines are complex information retrieval tools because they index documents worldwide and are used by users with diverse backgrounds. Due to such complexity, efforts have been made to understand how individuals look for information and what factors influence their behaviour (Kim, *et al.*, 2007). Thus, Web search engines are not only very widely used, but they are also increasingly studied (Spink *et al.*, 2002).

### 2.2.2 Web-Based Instruction Tools

Web-based instruction (WBI) has been described as '*a hypermedia based instructional programme which utilises the attributes and resources of the World Wide Web to create a meaningful learning environment where learning is fostered and supported*' (Khan, 1998, p.63). WBI tools have become a popular alternative to the traditional classroom teaching methods because these virtual learning spaces are far more accessible to a much wider range of learners (Khan, 2005). More and more educational establishments, as well as companies responsible for providing training, are adopting this technology and are offering online learning programmes. Furthermore, students expect educational and training establishments to employ the latest technologies to provide high quality instruction and 24/7 support (Khan, 2005). WBI tools provide flexible teaching and learning environments for students through the provision of non-linear learning (Pituch & Lee, 2006). Students have the ability and freedom to control their own learning (Farrell and Moore, 2000), for example, through the use of different navigational tools, such as a main menu, a hierarchical map or an alphabetical index. This ability to match their design with students' preferences is vital in ensuring the users can interact with the WBI tool in a manner that is effective, efficient and satisfying (Dillon and Zhu, 1997). Therefore, it is essential for designers to understand how students learn and what factors affect the ways in which they learn.

In summary, it is clear that the success of both types of application relies upon the ability of the application to meet the needs and preferences of each individual from a diverse user group. If users' needs are successfully met, they will have a more effective interaction with Web-based applications and complete their tasks with more efficiency. There are many different factors that influence the preferences of users, with human factors being perhaps the most basic distinction between users.

## 2.3. Human Factors

Research in the area of human factors has flourished in the past two to three decades, especially when the World Wide Web made computing technologies accessible to a wider audience. Human factors can be described as those individual human characteristics that can affect the way we interact with computer systems (Treu, 1994). There are many human factors, for example, age, gender, cognitive style or level of experience. Previous research has found that the latter three have been shown to significantly affect users' influence on preferences for the interface design of Web-based applications (Chen and Macredie, 2010). The following three sections will present a detailed explanation and analysis of each of these human factors.

### 2.3.1. Cognitive Style

Cognitive style is concerned with an individual's consistent approach to organising and representing information (Riding and Rayner, 1998). It is a relatively stable human trait as it is based upon physiological elements that are fixed or inherited and therefore cannot be acquired by teaching (Jonassen and Grabowski, 1993). Research in the 1960s and 1970s identified a multitude of different dimensions of cognitive style, which included, for example: divergent/convergent (Hudson, 1966), reflection/impulsivity (Kagan, 1966), adaptor/innovator (Kirton, 1976), holism/serialism (Pask, 1976), visualizer/verbalizer (Richardson, 1977), field dependent/independent (Witkin *et al.*, 1977). Among these, a large number of studies that investigate user preferences for Web-based applications have focussed on the dimensions that can be described as wholist/analytic (Riding and Sadler-Smith, 1992), which describes the tendency for individuals to process information either as an integrated whole or indiscrete parts of that whole (Graff, 2003). Perhaps the most prominent and well-established of these dimensions is Wikin's Field Dependence/Independence (Witkin, *et al.*, 1977). This dimension has been widely studied over the last three or four decades and has been proposed to share several similar characteristics with the lesser studied Holism/Serialism (Pask, 1979; 1988). The reasons for researching the Holist/Serialist dimension are threefold. Firstly, this dimension has been suggested to have conceptual links to the Field Dependent/Independent dimension. As previous studies (e.g. Chen and Macredie, 2004; Chen and Liu, 2008) have shown that the Field Dependence/Independence dimension has significant effects on learners' preferences for the design of WBI, it can be assumed that the Holist/Serialist dimension is also significant because of such links. Secondly, previous literature has paid less attention to the Holist/Serialist dimension. Thirdly, and perhaps most importantly, Ford (2000) argues that the Holist/Serialist dimension shows potential when adapting information systems to the individual needs of the learners. The next two sections will introduce these two cognitive styles.

#### 2.3.1.1 Field Dependence/Independence

Field Dependence refers to the '*degree to which a learner's perception or comprehension of information is affected by the surrounding perceptual or contextual field*' (Johanassen and Grabowski, 1993, p. 87). Within this dimension, users can be split into three categories: Field Dependent, Field Independent and Field Intermediate. Field Dependent individuals tend to perceive objects as a whole, whereas Field Independent individuals focus more on individual parts of the task

(Witkin, *et al.*, 1977). Furthermore, Field Independent individuals prefer an analytical approach and can extract relevant cues to complete their tasks, whereas Field Dependent individuals prefer a more passive approach and tend to focus on the most salient cues, regardless of relevancy (Witkin, *et al.*, 1977). In addition, Field Dependent individuals are more externally directed, whereas Field Independent individuals are more internally directed. Intermediate individuals are neither wholly Field Dependent nor wholly Field Independent, but their preferences tend to lie in between these two extremes. In this way, Intermediate users can have both Field Dependent and Field Independent characteristics. Table 2.1 provides a summary of the differences between Field Dependent (FD) and Field Independent (FI) individuals.

**Table 2.1. Differences between FD/FI characteristics (Witkin *et al.*, 1977)**

| Field Dependent (FD) | Field Independent (FI) |
|---|---|
| • Externally directed | • Internally directed |
| • Easily influenced by salient cues in the interface | • Influenced less by social reinforcement |
| • Focus on global experience and struggle with individual elements | • Focus on analytical experience, good at taking individual elements out of context |
| • Prefer to work in groups | • Prefer to work alone |
| • Find it difficult to process information and link with existing knowledge | • Process information with own structure from previous knowledge |
| • Demonstrate fewer reasoning skills | • Demonstrate greater reasoning skills |
| • More likely to accept ideas as presented to them | • More likely to accept ideas through strengthened analysis |

A number of tools have been developed to measure a users' level of Field Dependence. The Group Embedded Figures Test (GEFT; Witkin, 1976) and the Cognitive Styles Analysis (CSA; Riding, 1991). These are described in the next chapter, where their advantages and limitations are analysed in terms of this study.

The Field Dependence/Independence dimension is widely studied in the area of Web searching and Web-based instruction because it shows the extent to which a learner is able to restructure information using salient cues and field arrangement (Weller, Repman, & Rooze, 1994). In addition to the table above (Table 2.1), previous research shows that Field Dependent and Field Independent users have clearly distinctive characteristics, which result in them having different preferences when it comes to using Web-based applications. These differences can be broadly grouped into three categories: content organisation, information format and information search strategy.

***Content Organisation***

In terms of content organisation, Field Independent individuals prefer it if they are able to choose the order in which they can process the content, whereas Field Dependent individuals will become confused if presented with such a choice (Chen and Macredie, 2004). Palmquist and Kim (2000) also found that Field Dependent users preferred content that provided them with a well-structured route. Direct guidance aids the more passive Field Dependent users to avoid experiencing lack of comprehension (Chen, 2000). In addition, Field Independent users prefer to have a larger viewing window, more so than the Field Dependent users (Riding and Grimley, 1999).

***Navigational Strategy***

Many studies have found that cognitive style is influential in deciding users' preferred navigational strategy. For example, Liu and Reed (1995) conducted a study involving sixty-three international college students in a hypermedia assisted learning language environment. Their findings showed that Field Independent users tended to jump from point to point using links in the index, whilst Field Dependent users closely followed the menu sequence provided by the authors. Dufresne and Turcotte (1997) also found that Field Dependent individuals prefer to have a fixed path to navigate in a WBI program. In another study, sixty-five postgraduate students took part in a hypermedia-based tutorial (Ford and Chen, 2000). Similar findings were identified, as Field Independent users were found to make greater use of the index to find information, whereas Field Dependent users preferred to acquire a more global understanding from the map. Furthermore, Lee, Chen and Liu (2007) found that Field Independent users tended to use the main menu more than Field Dependent users. In addition, they also found that Field Independent users spent less time navigating the area than Field Dependent users. Lee *et al*. (2005) found that Field Dependent users preferred linear navigation, whilst Field Independent users preferred non-linear methods. This finding agrees with that of Frias-Martinez *et al*. (2007), who examined the effect of individual differences on users' behaviour in a Web-based digital library. They found that Field Dependent users frequently used the back and forward buttons, therefore following a linear style navigation pattern. In their comparative review, Chen and Macredie (2002) conclude that previous research suggests that navigational tools, such as the map, index or find function, should be provided for Field Independent users. On the other hand, a more structured, step-by-step approach should be provided for Field Dependent users that includes additional guidance in the form of menus or map structures.

***Information Seeking Strategies***

In addition to content organisation and navigational strategy, cognitive style has also been found to have a significant effect on users' information seeking strategies (Ford, Wood and Walsh, 1994). Field Independent users are thought to be more active in their information seeking strategies than Field Dependent through seeking out related links, such as using more references within the content (Lee, Chen and Liu, 2007). Kim (1997) also found that Field Independent users were more actively engaged in their searching tasks as they tended to use search engines, the find function and input URLs more frequently. In addition, Palmquist and Kim (2000) conducted an investigation involving 48 undergraduate students, categorising them into Field Dependent, Field Independent, novices and experts using the GEFT. Students were asked to perform factual and topical search tasks on a university website. Their findings indicated that Field Dependent users made more frequent use of embedded links. However, Chen and Liu (2008) found that Field Independent users browsed less pages than Field Dependent users. This suggests that Field Independent users have the ability to focus on the more relevant information applicable to their tasks, which could be because of their ability to be analytical and task orientated (Ford *et al*, 1994). Kim *et al*. (2004) compared the search strategies of fifty users from different cognitive style groups using Web search engines. Their results showed that the Field Dependent group significantly demonstrated more repeated search attempts and total search attempts than the Field Independent group. In addition, the Field Dependent group also made more use of search operators. Additionally, Ford *et al*. (2005) conducted a study involving two hundred and fifty postgraduate students. Their results indicated that Field Dependent users used a higher level of Boolean searching while Field Independent users tended to use a higher level of Best Matching in their search queries. Furthermore, Field Independent users were found to make more use of the advanced search functionalities (Frias-Martinez *et al*., 2007).

Despite the multitude of literature supporting the argument that cognitive styles have a significant effect on the way users interact with a system, there are a few studies that have failed to find a significant effect on behaviour. Shih and Gamon (2003) investigated this issue using the t-test and their findings suggested that Field Dependent and Field Independent individuals did not differ significantly in their learning patterns. Similarly, Huang *et al.* (2007) failed to find any significant differences between Field Dependent and Independent users.

#### 2.3.1.2 Holist/Serialist

The Holist/Serialist construct was originally conceived by Pask (1979) and classifies users into Holists or Serialists. Holists are global in their approach to information seeking and will look for an overall picture of any given topic before focussing on the detail of specific parts. On the other hand, Serialists are more analytical, requiring in-depth detail of individual parts of a topic before seeking out links to achieve a broader understanding of how everything is connected (Sternberg and Lubart, 1996). Holists seem to be able to multitask with efficiency, focusing on many different aspects from different areas at the same time, whereas Serialists work from the bottom up, concentrating on small parts one-by-one before synthesising the material to gain a global understanding (Jonassen and Grabowski, 1993). A summary of the main differences between Holists and Serialists can be seen in Table 2.2.

**Table 2.2. Differences between Holist and Serialist characteristics (Pask, 1979)**

| Holists | Serialists |
|---|---|
| • Take a global approach and create conceptual links between objects early on.<br>• Is able to move between theory and real world examples from the beginning.<br>• Broad focus: likes to have more than one thing on the go at the same time.<br>• Internally directed. | • Take an analytical approach, examining individual topics before forming conceptual links.<br>• Analyse theory or real world examples separately, only joining together if necessary.<br>• Narrow focus: prefers to focus on completing one task before moving onto the next.<br>• Externally directed. |

If we compare Table 2.1 and Table 2.2, it is clear that there are some basic similarities between the two cognitive style dimensions. More specifically, Field Dependent users and Holists show similarities, as do Field Independent users and Serialists. For example, both Field Dependent users and Holists prefer to take a global approach, whereas Field Independent users and Serialists prefer a local analytical approach. However, this is not universally true as there are some contradictions to this rule. For example, Field Independent users and Holist individuals have a tendency to be internally motivated. Thus, this suggests that although there are some similarities, the relationship is not strictly Field Dependent = Holist and Field Independent = Serialist.

As mentioned in the introduction to this section, cognitive style is thought of as a relatively stable trait and cannot therefore be ‘learned’ by an individual. As the aforementioned studies suggest, individuals with different cognitive styles will have different preferences when using Web-based applications. Therefore, these applications need to be designed in a way that is flexible enough to support users with all types of cognitive style. More specifically, even though Field Dependent users are thought to be like Holists, the differences between these may influence the effectiveness of interaction if the same interface is provided for both of these types of users. For example, Field Dependent users tend to be externally directed and will tend to follow the structure of the content presented to them (Witkin *et al.*, 1977).

Holists, on the other hand, tend to be more internally directed and prefer to explore the content in their own preferred manner (Pask, 1979). Following these characteristics, if a suggested route is imposed upon a Holist or a Field Dependent user is left to explore content at their will, this could leave users feeling frustrated or confused and thus reducing the effectiveness of the Web-based application. Therefore, further research into the differences between these dimensions of cognitive style would be advantageous for a number of reasons. Firstly, with regards to practical applications, such research would help designers to create more effective Web-based applications. Secondly, in regards to theory, this type of research would help add to the body of knowledge of user characteristics that belong to each of these cognitive styles. Thirdly, as the Holism/Serialism dimension is less studied than the comparatively more popular Field Dependence/Independence dimension, studying the relationship between these would enable researchers to overcome this gap in research and perhaps begin to infer characteristics from the latter to the former.

### 2.3.2 Prior Knowledge

As well as cognitive style, a user's level of prior knowledge has been shown to influence their preferences for interaction in a Web-based application (Gauss and Urbas, 2003; Ford and Chen, 2000; Reed, Oughton, Ayersman, Giessler and Ervin, 2000). Prior knowledge is made up of both system experience and domain knowledge. The former refers to users' knowledge of the system being used whereas the latter refers to users' understanding of the content area. As Web-based applications are online and involve elements of both computing and the Internet, previous research demonstrates that system experience, which in this case specifically refers to users' experience with using computer and Internet based systems (Lazonder, 2000). In addition, it has a particularly influential effect on students' learning (Minetou, Chen and Liu, 2008).

The level of system experience is not necessarily shown to impact upon students' learning outcomes. However, it has been shown to affect the way that students go about organising their tasks (Gauss and Urbas, 2003). In their study, Shih, Munoz and Sanchez (2006) analysed one hundred and twenty undergraduate students who participated in a Web-based course that was designed to train personnel trainers to apply ICT in training. They found that users with more experience took less time and visited fewer pages per work session to complete their tasks when compared to those with less experience. Similarly, Lazonder *et al.* (2000) used twenty five pre-university students to examine the differences between novices, those users with low levels of previous experience, and experts, those users with high levels of previous experience. They found that users with higher levels of Web experience performed their tasks quicker, more accurately and in fewer actions than those with lower levels of experience. Furthermore, Ford and Chen (2000) found that users with higher levels of experience could reach a more detailed level of subject content than those with lower levels of experience. Hölscher and Strube (2000) found that users with lower levels of experience were less flexible in their searching strategies, preferring to return to previous stages in their search rather than attempting a new approach. In addition, users with high levels of knowledge have been shown to perform searches that are content driven and use technical terms (Marchionini, 1998). Jenkins, Corritore and Wiedenbeck (2003) investigated the different search strategies employed by novices and experts when searching for medical data using search engines. They found that experts used depth-first strategies, following links one after another until detailed information was located. Novices, on the other hand, employed breadth-first strategies, using the hierarchical menu to build up an overview of the content without exploring content in any great detail.

In addition to task organisation, previous experience has been shown to have an effect on the way that users approach their tasks. For example, eighteen graduate students took part in a study by Reed, Oughton, Ayersman, Giessler and Ervin (2000) that aimed to identify the extent to which prior computer-related experiences affected navigational patterns of students in a learning environment. They found that a more linear approach to navigation was taken by those that had lower levels of previous experience. Comparatively, those that had higher levels of previous system experience took a more non-linear approach.

Table 2.3 summarises the differences between novices and expert characteristics. These studies suggest that students with higher levels of previous system experience have the ability to organise their tasks in a more efficient manner and are able to navigate their environment in a way that suits their learning style more effectively than those with lower levels of system experience. It is therefore important to identify where the weaknesses lay in those students with lower levels of system experience so that interface features can be customised to support them.

**Table 2.3. Differences Between Novice and Expert Characteristics**

| Novice | Expert |
|---|---|
| • Local mental models<br>• Undirected search (trial and error)<br>• Surface features<br>• Mental simulation of isolated functions<br>• Incomplete analysis<br>• Breadth-first strategies<br>• Design pieces<br>• Failure to integrate pieces into a whole<br>• Find a (any) solution | • Global mental models<br>• Directed search<br>• Deep structures<br>• Mental simulation of integrated functions and whole application<br>• Complete analysis deferring details<br>• Depth-first strategies<br>• Design whole and add pieces<br>• Integrated whole throughout the process<br>• Find the best solution |

* Minetou, Chen and Liu (2008), adapted from Chen, Fan, and Macredie (2006)

However, the term ‘system experience’ used in the aforementioned studies includes both computer and Internet related experiences. The former is a general term that describes knowledge of any computer related experiences, whereas the latter is more specific and covers knowledge of the Internet and the things an individual can do using the Internet (Potoksky, 2007). Internet experience is concerned specifically with previous Internet usage, e.g. browsing the Web whereas computer experience is a more general term that can encompass anything from burning a DVD to setting up networks (Schumacher and Morohan-Martin, 2001). Up until recently, access to the Internet was through a keyboard and computer monitor (Durndell and Haag, 2002) and it was also necessary to have the knowledge of how to connect to networks. Therefore, it was more common to find users with more computer experience than Internet experience. This is due to the fact that users with lesser computer experience not only face difficulties in learning how to set up a keyboard and monitor, but they found the technical barriers associated with connecting to the network hard to overcome (Kraut, *et al.*, 1996). However, the widespread availability of home computing and more recent technological advances mean that users can now connect to the Internet without as much technical know-how, and as such it is becoming increasingly common to find users with equal or more Internet experience than computer experience.

Regardless of whether currently students have more Internet experience than computer experience, it seems that these two types of experience involve different skill sets. Some research has suggested that these two types of knowledge are distinctive from each other (Bradlow *et al*., 2002), even though Internet experience is technically a type of computer experience. As a consequence, the majority of tools designed to identify users' previous experience include both Internet experience and computer experience. Furthermore, they only roughly consider the differences between novices or experts, instead of the different types of experience and associated characteristics that they have. Such a mixture may cause students to feel frustrated when using WBI tools because the design of these tools is not matched with the computing or Internet skills that they specifically have. Therefore, it is necessary to conduct further studies to identify the preferences of students with differing types of system experience, i.e. Internet experience and computer experience. Furthermore, this information will help to determine if computer experience and Internet experience can really be categorised together or whether they should be separated and treated as different factors. The benefit of doing this will help designers to provide the necessary level of support for students with differing levels of computer experience and Internet experience, and thus aid in the development of more effective Web-based application tools for the future and enabling students to be more confident and successful in their learning.

### 2.3.3 Gender Differences

Among the many individual differences, gender is perhaps the most basic and obvious difference between users. In the past, there has been a great divide between male and female usage of technology. Morahan-Martin (1998) states that, from childhood to adulthood, males are more likely than females to use and have more favourable attitudes towards computers. In addition, previous research often found that males tend to have more computer experience, use many different computer applications and voluntarily choose to use the computer more than females (Modianos & Hartman, 1990; Morahan-Martin *et al*., 1992). Nowadays, this gap is rapidly disappearing as female presence increases in a multitude of professions (Moss *et al.*, 2006) and women are now more likely to use the Internet at work (Kominski & Newburger, 1999). Indeed, Katz, Rice & Aspden (2001) report that women accounted for the majority of new Internet users during the period of 1997-2000. More recent research has found that males and females have been shown to use computers and the Internet for different reasons. For example, males tend to use the Internet mainly for entertainment and leisure, whereas women tend to use it for interpersonal communication and for educational purposes (Weiser, 2000). Jackson *et al*. (2001) also found that women tend to use the Internet to communicate, whereas males tend to use it more to search for information.

Despite prominent physiological differences between the sexes, Hyde (2005) argues that males and females generally show similar abilities on the majority of psychological variables. However, males and females have been shown to process information in very different ways. Meyers-Levy (1988) conducted a study that examined the effect of gender differences on information search behaviour. They found that males tended to process only highly available information and relied upon their own opinions to make faster decisions. On the other hand, females followed a more exhaustive search pattern and relied more on a broad variety of information from a wide range of external sources. This is echoed in the findings of Riding and Rayner (1998), where males were shown to process information to a superficial level, whilst females processed information to a much deeper level. Such different approaches to information processing mean that the two sexes will have different preferences for information searching tools when using Web-based applications. However, past research into the effects of gender on users' preferences of Web-based applications presents many conflicting and inconclusive results. The majority of previous research follows with the general idea that females experience more anxiety and thus encounter more problems whilst interacting with the application than males, who are more confident in such an environment (e.g. Gunnmales, McSporran, MacLeod & French, 2003; Morahan-Martin, 1998; Ford & Miller, 1996). For example, Shashanni (1994) found females reported more feelings of fear and helplessness when interacting with computers, whilst males showed more interest in learning about computers. Furthermore, it has been shown that males are more likely to attribute failure to external factors, whilst women are shown to look initially to internal factors (Lufkin & Wiberg, 2007). In addition, many studies have found that men and women show different preferences in terms of layout and presentation. For example, female users prefer the layout to uses lighter colours, whilst male users prefer darker colours, i.e. black and blue (Moss *et al*. 2006). Another study found that boys preferred green and blue, whereas girls preferred red and yellow (Passig & Levin, 1999).

In terms of content organisation, it has been found that gender is influential in deciding users' navigational patterns (Roy, Taylor & Chi, 2003). Several studies have found that a simple navigational system that is easy to use is very important to female users (e.g. Oser, 2003). Males tend to navigate in a much broader and non-linear approach than females (Large *et al*, 2002). Furthermore, these authors also discovered that male users also tend to be more actively involved than female users, as they had more page jumps and spent less time viewing pages. Table 2.4 summarises the differences in characteristics between males and females.

**Table 2.4: Differences between male and female characteristics**

| Male | Female |
| --- | --- |
| • Broad, non-linear navigational approach<br>• Rapid search strategies<br>• Active<br>• High confidence, blame external factors for errors. | • Narrow, linear navigational approach<br>• Exhaustive search strategies<br>• Passive<br>• Low confidence, blame internal factors for errors. |

### 2.3.4 Cognitive Style vs. Prior Knowledge vs. Gender Differences

In summary, it is clear that the three human factors, cognitive style, system experience and gender differences, have a significantly influential effect on users' preferences of Web-based applications. The previous three sections also identify links between different dimensions of cognitive styles (i.e. Field Dependence/Independence and Holism/Serialism, and Internet experience and computer experience). In addition to this, previous research has also highlighted the interaction between these three human factors. For example, Novices, Females and Field Dependent users are purported to display similar behaviour and preferences (Chen, 2000). For instance, Field Dependent users are externally directed, preferring to be given a suggested route through content (Witkin *et al*., 1977). Likewise, novice users lack previous experience in an environment and therefore prefer to be given a suggested route to help them orientate themselves. Females are generally less confident about using technology (Abbott & Bienvenue, 2007) so therefore they perform more efficiently when they have been given a suggested route. Likewise, this is purportedly the same for Experts, Males, and Field Independent users (Fan, 2005). For example, Field Independent users, experts and males all have been shown to prefer non-linear approach to navigation (Lee *et al*., 2005; Large *et al*., 2002; Reed, *et al*., 2000). Despite these similarities, there are, however, some differences. For example, both Field Dependent users and males are generally more holistic in their approach in that they prefer to attain an overview early on and then process information at a more detailed level (Witkin *et al*., 1977; Riding, 1998). Expert users, on the other hand, prefer more of a depth-first approach (Jenkins, Corritore & Wiedenbeck, 2003), exploring each topic in detail before moving on to the next. These differences show that just because these types of user exhibit similar behaviours, it is not always correct to say that they are totally equal. Therefore, investigations that aim to identify the extent of the relationships between these factors can help to identify the needs of each type of user and thus help designers to increase the effectiveness of Web-based applications.

The next section will describe the methods of analysis that have been and are currently used in studies that investigate the relationship between human factors and users' preferences for Web-based applications.

## 2.4 Analysis of Human Factors

As is evidence in the previous section, a multitude of studies have investigated the influence of cognitive style, prior knowledge and gender differences on users' preferences for Web-based applications. The majority of these studies employ statistical tests to analyse relationships and find correlations between human factors. As can be seen from Table 2.5, which presents a sample of these studies, popular statistical methods include ANOVA or T-Tests (e.g. Mitchell, Chen & Macredie, 2005), Pearson's Correlation (e.g. Chen, 2005), Multiple Regression (e.g. Ford, Miller & Moss, 2001) and Factor Analysis (e.g. Chuang and Tsai, 2005). Although many meaningful results have been found regarding the influence of a single human factor on users' preferences for Web-based applications, these statistical tests seem to lack the power to perform when comparing multiple human factors.

There can be a number of possible reasons for this. Firstly, these techniques lack methods to control the quality of the data. In other words, if the quality of the data input into the analysis is low, the results will not accurately represent users' true preferences. In particular, human preferences data is by nature very 'noisy', meaning it will often include data items that can be considered outliers. Consequently, the inclusion of these data items in the statistical analyses can introduce bias in the results achieved through such techniques. These data items might be colouring the results in a way that users' true preferences are hidden and thus might explain why some studies (e.g. Huang, *et al.,* 2007) fail to find correlations between human factors or dimensions within human factors.

Another possible reason lies within the requirement to make assumptions before running some of these tests, for example, defining dependent and independent variables. The chances of discovering relationships between the data are slim if they are neglected in the process of making assumptions (Lee, *et al.*, 2009).

Furthermore, these statistical techniques do not have the ability to sufficiently capture relationships between different variables. As is often the case with human factors, it is the interaction of one or more factors that affect user preferences. Therefore, a method is needed that a) is able to filter out the noisy data, b) overcome

**Table 2.5. Statistical Methods of Analysis in Previous Studies**

| Authors | Web Applications | | Human Factors | | | Method of Analysis |
|---|---|---|---|---|---|---|
| | **Web Search Tools** | **Web-Based Instruction Tools** | **Cognitive Style** | **Prior Knowledge** | **Gender Differences** | |
| Chen (2005) | | X | X | X | X | Pearson's Correlation |
| Chen, Magoulas and Dimakopoulos (2004) | X | | X | | | Frequency tables |
| Chuang and Tsai (2005) | | X | | X | X | Factor Analysis |
| Cutmore, *et al*. (2000) | | X | X | | X | ANOVA |
| Ellis, Ford and Wood (1992) | | X | X | | | Chi Squared, Pearson's, ANOVA |
| Ford and Chen (2001) | | X | X | | X | ANOVA/T-Tests |
| Ford, *et al*. (2002) | X | | X | | | Spearman test |
| Ford, Miller and Moss (2001) | X | | X | X | X | Multiple Regression, Factor Analysis |
| Graff (2003) | | X | X | | | MANOVA |
| Lu, *et al*. (2003) | | X | X | | | Frequency, Percentage, Pearson's Correlation |
| Mitchell, Chen and Macredie (2005) | | X | | X | | ANOVA, Pearson's Correlation |

the need to assume relationships within the data and c) sufficiently model the interrelationships between variables.

Researchers have attempted to overcome this problem by introducing Data Mining into the analysis of human factors and users' preferences. Data mining is the search for valuable information within large volumes of data (Hand, Mannila & Smyth, 2001) by systematically exploring underlying patterns, trends, and relationships hidden in available data (Lee, *et al.*, 2006). Unlike traditional statistical methods that try to prove a known relationship (Moss & Atre, 2003), data mining techniques use the data itself to predict, model or identify interrelationships within the data (Urtubia, *et al.*, 2007).

Data mining methods can generally be grouped into four categories: classification, clustering, association rules and information visualisation. The following sub-sections will describe these in further detail.

### 2.4.1 Classification

Classification is a widely used technique for extracting meaningful information from a dataset (Baglioni *et al.*, 2005). Classification uses algorithms to find a model that describes a data class or concept (Han & Kamber, 2006). By identifying a series of predefined labels, items can be categorised into classes according to their attributes (e.g. 'age' or 'gender'). Thus, its purpose is to predict the label of a new item based on a set of already labelled items (Szpunar-Huk, 2006). For example, in the case of a bank loan clerk, classification is useful for predicting whether loan applicants are a 'safe' or 'risky' investment for the bank based on the class that they belong to. In this way, classification techniques can be used to gain a better understanding of the existing data and to predict how new classes will behave (Liu & Kellam, 2003).

Decision trees are one of the most frequently used classification techniques, employed to discover rules and relationships by systematically dividing information contained within data (Chen, Hsu & Chou, 2003). They classify data by constructing tree-like structures through a series of Boolean functions, i.e. "yes" or "no" questions based on the characteristics of a set of variables, until no more relevant branches can be derived. Popular decision tree algorithms include Chi-squared Automatic Interaction Detection (CHAID) (Kass, 1980), Classification and Regression Trees (CART), (Breiman *et al.*, 1984), and C4.5 (Quinlan, 1993). Amongst these algorithms, there are some differences, one of which is the capability of modelling different types of data. All of these algorithms can support the modelling of categorical data whilst only the C4.5 and the CART can be used for the modelling of numerical data. The other difference is the process of model development, especially at the stages of tree growing and tree pruning. In terms of the former, the C4.5 splits a tree model into as many ramifications as necessary whereas the CART algorithm can only support binary splits. Regarding the latter, the pruning mechanisms located within the C4.5 and CART support the removal of insignificant nodes and ramifications while the CHAID hinders the tree growing process before the training data is being overused.

### 2.4.2 Clustering

Classification is thought of as a supervised learning technique because it uses a set of predefined class labels. However, when the classes are unknown in the data, an unsupervised learning technique such as clustering can be applied. In this way, clustering can uncover previously hidden and unexpected trends or patterns in data because no assumptions are made regarding the structure of the data. Clustering involves grouping items into 'natural' clusters based on their similarity to one another (Hand, Mannila & Smyth, 2001). Each item in a cluster is similar to those within its cluster, but dissimilar to those in other clusters (Roussinov & Zhao, 2003). Clustering is commonly used to discover groupings and populations within data. For example, clustering techniques can be used to identify customer affinity groups in order to begin to build customer profiles.

### 2.4.3 Association Rules

Association rules were first introduced by Agrawal, Imielinski, & Swami (1993) and can be defined as finding rules in transaction data that satisfy the minimum support and confidence constraints. Today, association rules are mainly used to find relationships between two or more items or to find events that occur at the same time in databases. These rules are normally expressed in the form (X→Y), where X and Y are items. In any set of transactions, this means that those transactions containing the items X, generally tend to contain the items Y. A rule is usually measured by support and confidence, where the support is the percentage of both X and Y contained in all transactions and the confidence is calculated by dividing the number of transactions supporting the rule by the number of transactions supporting the rule body (Zhang, Gong & Kawamura, 2004).

### 2.4.4 Information Visualisation

Information visualisation is based on the core assumption that human beings perceive structure better in visual formats and uses visual aids to help understand these key relationships. The basic idea is to present the data with some graphics, e.g. 2D and 3D graphics, allowing the human to gain insight from the data, draw conclusions, and directly interact with the data (Ankerst, 2001). Since the user is directly involved in the exploration process, shifting and adjusting the exploration goals is automatically done if necessary (Lopez, Kreuseler & Schumann, 2002). This approach is especially useful when little is known about the data and the exploration goals are vague. There are several well-known tools for visualization, which can be divided into: standard 2D/3D displays, dense pixel displays, geometrically transformed displays, icon-based displays, and stacked displays. These tools differ in how they arrange the data on the screen and how they deal with multiple dimensions in case of multidimensional data (Keim, 2002). No single visualisation tool will be suitable to address all problems so different tools must be chosen based on the task and data. However, easy access to external databases and a sufficient number of analytical methods are two important criteria for selection of suitable tools.

In summary, it is clear that these four types of data mining techniques could be used to identify the effect of human factors on user preferences. However, classification appears to be the most appropriate for a number of reasons. Firstly, the class labels are already known for each of the human factors that will be analysed. These include: Field Dependent, Field Independent, Holist, Serialist, Novice, Expert, Male and Female. Secondly, classification techniques allow the accuracy of the classification results to be measured, thus providing a way to select the most accurate result to represent users' preferences. Thirdly, decision tree classification provides a visual method that can represent the results in a way that can be easily interpreted. Finally, decision tree classification has been used successfully in a number of previous studies to identify the effect of human factors on users' preferences.

For example, a study involving 129 undergraduate business students constructed a recommendation system using decision trees (Zhu, Greiner, and Haubl, 2003). This recommendation system was based upon annotated weblog data and helped the users to find information relevant to their search on the Web. Similarly, Liu and Kešelj (2007) used decision trees to automatically classify the navigational patterns of users to predict the next pages that were more likely to be visited next. Furthermore, Lee, *et al*., (2009) used decision trees to classify users' preferences for a Web-based learning tool. The aim of their study was to discover how cognitive style influenced users' navigational behaviour.

In summary, decision tree classification would indeed provide a solution that covers two out of the original three goals for an ideal method for analysing the relationships between human factors and their influence on users' preferences of Web-based Applications. However, the performance of the classification would still be impaired by the inclusion of noisy or irrelevant data. Therefore, it is necessary to use an additional technique that would help remove such items from the dataset. Feature selection is a method that is popular for solving this problem and is employed used to minimise the effect of those items, or features, that are irrelevant by selecting relevant subsets from the original dataset (de Souza, Matwin and Kapkowicz, 2006). The following section introduces the feature selection method.

### 2.4.5 Feature Selection

Generally, feature selection techniques can be divided into two categories: filters, which produce a ranking of all features without involving any classifiers; and wrappers, which use classifiers to evaluate subsets and interactions of features (Yang and Olafsson, 2006). Although wrappers usually provide better performance (Raman and Ioerger, 2002), traditionally wrapper methods consider just one classifier. The problem with this method is that each classifier will have its own biases. Thus, each classifier will select a different feature subset which may lead to varying levels of accuracy. Therefore, there is a need to consider multiple classifiers, instead of just the one classifier as is traditionally used. Identifying the features commonly selected by several classifiers could maximise the overall effectiveness of feature selection by making sure that only the most relevant subset of features is chosen (Chrysostomou, et al. 2008). Once this relevant subset of features is chosen, a classification technique can be used to illustrate the relationships between the relevant features and a particular target variable (e.g., cognitive styles). Among various classification techniques, decision trees can be used to identify the accuracy of the relevant feature sets. This method, which is otherwise known as wrapper-based decision trees (WDT), was developed by Chrysostomou (2008). The author used multiple classifiers from four different families, including Bayesian Networks, Decision Trees, Nearest Neighbour and Support Vector Machines, to select relevant features from two HCI datasets. The resulting feature sets were then classified using decision trees with the aim of identifying how the number and nature of classifiers influenced the selected feature sets. The results showed that using few classifiers resulted in many relevant features being selected, whereas using many classifiers resulted in few relevant features being selected. In addition, it was found that using three classifiers resulted in highly accurate feature sets. Regarding the nature of classifiers, Decision Tree, Bayesian Network and Nearest Neighbour classifiers were shown to have a very significant influence on the number and accuracy levels of features selected.

The advantages of such a method are threefold. Firstly, using decision trees allows the relationships among the features to be visually demonstrated and easily interpreted. Secondly, there is more chance of overcoming any bias if multiple classifiers from different classifier families are used in the feature selection stage. Therefore, only highly relevant feature subsets will be selected. Finally, by combining highly relevant features with an accurate decision tree, more reliable relationships can be identified.

### 2.5 Summary

This chapter presented a review of the previous literature investigating the effect of human factors on users' preferences for Web-based applications. Three human factors were identified, including cognitive style, system experience and gender differences. The review demonstrates that there is a need to conduct further research into the key factors that significantly affect users' preferences for Web-based applications. More specifically, cognitive style, system experience and gender differences were reviewed and a number of significant links were identified. In addition, suitable data mining tools were identified for analysis.

# Chapter 3 – Study 1: Field Dependent/Independent vs. Holist/Serialist

## 3.1 Introduction

This chapter will describe the first of the three studies examined within this thesis and will focus in particular on cognitive style and its effect on users' preferences of Web-based applications. Specifically, two dimensions of cognitive style are analysed: Witkin's (1976) Field Dependence/Independence and Pask's (1979) Holist/Serialist. As discussed in the previous chapter, the first of these two dimensions is very popular in previous studies, although the latter lacks the number and quality of previous works. Some researchers (e.g., Ash, 1986; Brumby, 1982; Entwistle, 1981; Jonassen and Grabowski, 1993; Riding and Cheema, 1991) have even proposed a link between the two, with the idea that Field Dependent individuals have similar preferences to Holists and Field Independent individuals have similar preferences to Serialists. Therefore, the aim of the study presented in this chapter is to examine the nature of this link. More specifically, the following research question is proposed as the goal for this particular study:

> *RQ1: To what extent is the similarity between Witkin's Field Dependent/Independent and Pask's Holist/Serialist dimensions of cognitive style?*

Such an examination will make contributions to the field in a number of ways. Firstly, this will help to gain a deeper understanding of the preferences of users belonging to different cognitive style groups. Secondly, if it is indeed found that there is the hypothesised relationship between the two cognitive styles, we can begin to infer known characteristics from the well established Field Dependent/ Independent dimension onto the less studied Holist/Serialist dimension. This would be of invaluable use to researchers and designers concerned with the Holist/Serialist dimension. Figure 3.1 presents the framework of this study. The following sections describe the procedure used to collect the data and then the data mining framework used to analyse the data. The findings will then be discussed and any conclusions will be drawn.

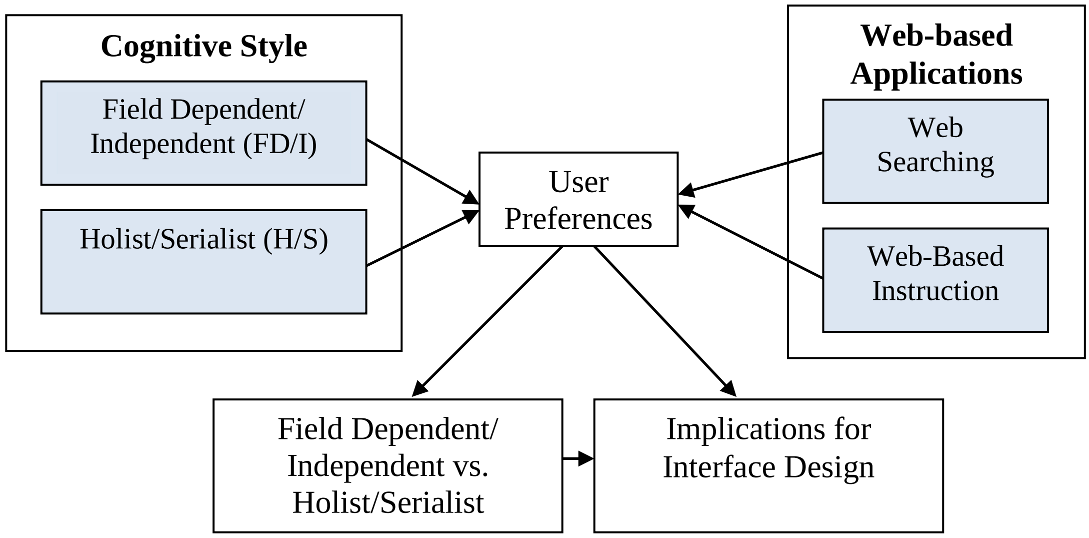



**Figure 3.1: The framework of Study 1**

### 3.2 Data Collection: Dataset 1 – Web Search Tools

As discussed in the previous chapter, two Web-based Applications were studied throughout this thesis, including Web search tools and Web-based Instruction tools. The following sections describe the procedures involved in collecting the data related to Web Search Engines (Dataset 1), covering the participants involved (3.2.1.), the research instruments used (3.2.2.) and the experimental procedure (3.2.3.).

#### 3.2.1 Participants

120 students from a UK university, who all had the basic computing and Internet skills necessary to use Web search engines, participated in this study. According to the results of the exit questionnaire (See Section 3.2.2.3), the participants were almost evenly divided among the sexes (61 males, 59 females). Likewise, the participants' cognitive styles were roughly evenly distributed, in terms of both Field Dependence/ Independence and Holists/Serialists. More specifically, there were 40 Field Dependent users, 33 Field Intermediate users and 47 Field Intermediate users and there were 63 Holists and 57 Serialists.

### 3.2.2 Research Instruments

The research instruments used in this study included the Cognitive Style Analysis (CSA), the Study Preference Questionnaire (SPQ) and an exit questionnaire. The CSA classified participants as either Field Dependence or Field Independent, whereas the SPQ identified whether participants were Holists or Serialists. Additionally, an exit questionnaire was used to identify users' preferences for interface features of the examined search engines, i.e. Google and Yahoo. The following sections introduce and explain the three types of instrument.

#### 3.2.2.1 Cognitive Styles Analysis (CSA)

The CSA was developed after Riding and Cheema (1991) identified that the majority of cognitive styles fell into two broad categories: a verbal-imagery dimension and a wholistic-analytic dimension. The CSA includes three sub-tests, the first related to assessing the verbal-imagery ratio of a user and the second two are related to assessing the wholistic-analytic ratio. This study focuses on the latter of these two dimensions (i.e. Field Dependence/Independence, Holist/Serialist), so only the latter two tests are described here. The first of these two wholistic-analytic sub-tests presents items containing pairs of complex geometrical figures that the individual is required to judge as either the same or different. The second presents several items, each comprising of a simple geometrical shape, such as a square or a triangle and a complex geometrical figure. The individual is asked to indicate whether or not the simple shape is contained in a complex one by pressing one of the two marked response keys (Riding and Grimley, 1999). These two sub-tests have different purposes. The first sub-test is a task requiring the Field Dependent capacity, whereas the second sub-test requires the disembedding capacity associated with Field Independence. In this way, Field Dependent competence is positively measured rather than being inferred from poor Field Independent capability (Ford and Chen, 2001) as it is in other FD/I tests, for example, the Group Embedded Figures Test (GEFT). This study follows Riding's (1991) recommendations that scores below 1.03 denote Field Dependent individuals; scores of 1.36 and above denote Field Independent individuals; and scores between 1.03 and 1.35 are classified as Intermediate.

Numerous papers have been published describing studies involving the CSA, giving it a strong background of theoretical support. In addition, many researchers have investigated the validity of the CSA (e.g. Riding and Cheema, 1991; Riding and Grimley, 1999; Riding and Rayner, 1998), with most of these studies providing evidence of construct validity (Rezaei and Katz, 2004). However, it must be noted that conflicting views are held by some researchers who have questioned the validity of the instrument (e.g. Peterson, Deary and Austin, 2007; Rezaei & Katz, 2004). These authors claim that the CSA test, in particular the verbal-imagery dimension, is unreliable and not internally consistent. Despite this controversy, the CSA is the most frequently used computerised measure of cognitive style in UK and European institutions (Rezaei and Katz, 2003).

#### 3.2.2.2 Study Preference Questionnaire (SPQ)

SPQ is an 18-item inventory for assessing the strategies of Holists or Serialists. Users are provided with two sets of statements, one on the right and the other on the left. They were asked to indicate their degree of agreement with either statement, or to indicate no preference (Ford, 1985). This study will identify Holists and Serialists by using the following criteria: (a) if users agree with over half of statements related to Holists, they are treated as Holists; and (b) if users agree with over half of statements related to Serialists, they are treated as Serialists. The SPQ was chosen because it has been used in several previous studies (e.g. Ford and Chen, 2000; 2001). However, it must be noted here that its reliability remains unstudied and there are still no published norms (Ford *et al*., 2002).

As shown in the above two sections, the CSA classifies users into three categories: Field Dependent (FD), Field Intermediate and Field Independent (FI); whereas the SPQ classifies users into two: Holists (H) or Serialists (S). It is noticeable here that there is an imbalance in the number of categories involved in the comparison. The Field Intermediate category is the ‘odd one out’ as is formed of individuals whose tendencies are neither wholly Field Dependent nor wholly Field Independent. Instead, they have a mixture of characteristics from both tendencies. As a result, it is very hard to correctly classify Field Intermediate behaviours, which can often cause confusion in the results. For example, in a study by Lee, Chen and Liu (2007), the majority of students who were misclassified in the decision tree were Field Intermediate. Therefore, there is a need to reconsider this category of user. As indicated in Section 3.2.1, the participants were identified as FD, Intermediate or FI based on the scores obtained from the CSA. The scores identified for the Intermediate group are between Field Dependent and Field Independent. Thus, the Intermediates were re-allocated based on their mean score (i.e., 1.20). More specifically, those users that scored below this mean were re-classified as Field Dependent (FD) as their behaviours were closer to this group, whilst those that scored above this mean were re-classified as Field Independent (FI). The new distributions of participants' cognitive styles can be seen below in Figure 3.2.

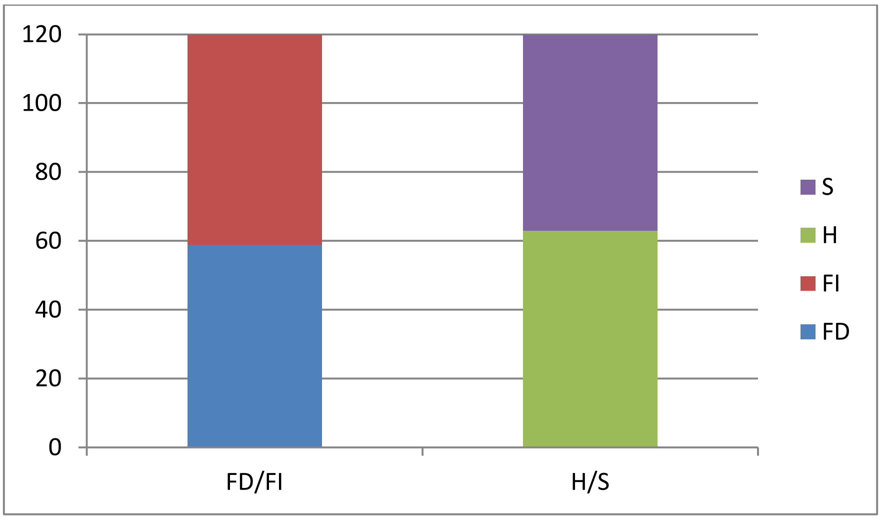



**Figure 3.2. Dataset 1 - Participants' Cognitive Style Distributions**

#### 3.2.2.3 Exit Questionnaire

The exit questionnaire for Dataset 1 consisted of 90 questions, which were designed to identify users' preferences for the interface design of search engines. These questions were developed based upon the well-established Neilson's Usability Heuristics (Nielson, 2001). These heuristics have been validated and adapted to study many Web-based applications (Nielson and Norman, 2000) and ensure a holistic approach to evaluating an interface to make sure that a range of issues within the problem space are covered (Chen and Macredie, 2005). The questions particularly emphasised the interface elements that were thought to be the most problematic for users when interacting with the system. A sample of these questions can be found above in Table 3.1 (the full questionnaire is listed in Appendix A).

**Table 3.1. Examples of the questions in the questionnaire**

| Section | Questions |
|---|---|
| Visibility | • There is visual feedback when objects are selected.<br>• The names of links and buttons are understandable.<br>• You can easily identify available options. |
| Consistency | • Navigation buttons are consistent from one screen to another.<br>• Vertical and horizontal scrolling is possible in each window.<br>• Consistent colour scheme is used in the system. |
| Error Messages | • Error messages let you know the cause of the problem.<br>• The error that you have made is highlighted.<br>• Error messages suggest what action you need to take to correct the error. |
| Flexibility | • The results can be sorted in a variety of ways.<br>• There is a detailed alphabetical index to help you locate specific information.<br>• A variety of search options are available to you. |
| Online Help | • The help provides you with step-by-step instructions.<br>• There is an alphabetical index to help you to locate information within the help system.<br>• The help can provide you with enough information to illustrate common errors. |

Participants were required to respond to all the questions using a five-point Likert scale, including ‘very unimportant’, ‘unimportant’, ‘neutral’, ‘important’ and ‘very important’. Questions were also added to gather demographic information on the participants, identifying their gender and the levels of their previous system experience in using computers and the Internet. It took approximately 15 minutes to fill out the whole questionnaire.

### 3.2.3 Experimental Procedure

The experimental procedure to collect the Web search engines dataset involved participants taking part in a three-phase study described subsequently. Firstly, the participants took the CSA (as described in Section 3.2.2.1) and the SPQ (as described in Section 3.2.2.2) to determine their cognitive styles, one for Field Dependence/ Independence and one for Holism/Serialism.

After identifying their cognitive styles, participants were asked to interact with two Web search engines for the second phase. The two search engines used in this study were Google and Yahoo, which are two of the most popular used to date. When interacting with these two search engines, the participants were asked to perform twenty tasks, including both browsing tasks and searching tasks, that provided them with the opportunity to experience the range of features offered by the search engine. The browsing tasks involved participants being asked to find information by browsing the categories within the search engine (e.g. Google Directory), whilst participants used only the main search box for the searching tasks. The tasks were designed so that each had only one right answer and covered a number of topics (e.g. education, literature and geography) to avoid bias to participants with particular subject knowledge. The tasks were evenly divided between the two search engines and the tasks performed on each search engine were comparable to the other in terms of type and scope. For example, if the participant was asked to browse for '*The name of the Mayor of the City of Leeds*" in Google, the participant would then be asked to browse for '*The name of the Mayor of the City of Sheffield*" in Yahoo. Likewise, if the participant was asked to search for '*The Chief Editor of the International Journal of Information Management*" in Google, the participant would then be asked to search for '*The Chief Editor of the International Journal of Human-Computer Studies*" in Yahoo. Participants were given an hour to complete the browsing and searching tasks.

Upon completion of these tasks, participants were asked to fill out the exit questionnaire (as described in Section 3.2.2.3). The results of this questionnaire were collected and then taken on to the data analysis stage.

## 3.3 Data Collection: Dataset 2 – Web-Based Instructional Tools

The second dataset was related to users' preferences of Web-based Instruction tools. The following sections describe the procedure involved in collecting the data.

### 3.3.1. Participants

This dataset involved 65 postgraduate students from a UK university. A request was issued to students in lectures and by e-mail, making clear the nature of the study and their participation. Again, the sample was evenly divided in terms of gender (32 males, 33 females) and all volunteers had the basic computing and Internet skills necessary to operate a WBI tool. The participants' cognitive styles were roughly evenly distributed, in terms of both Field Dependence/Independence and Holists/Serialists (Figure 3.3). More specifically, there were 29 Field Dependent (FD) users, 36 Field Independent (FI) users, and 33 Holists (H) and 32 Serialists (S).

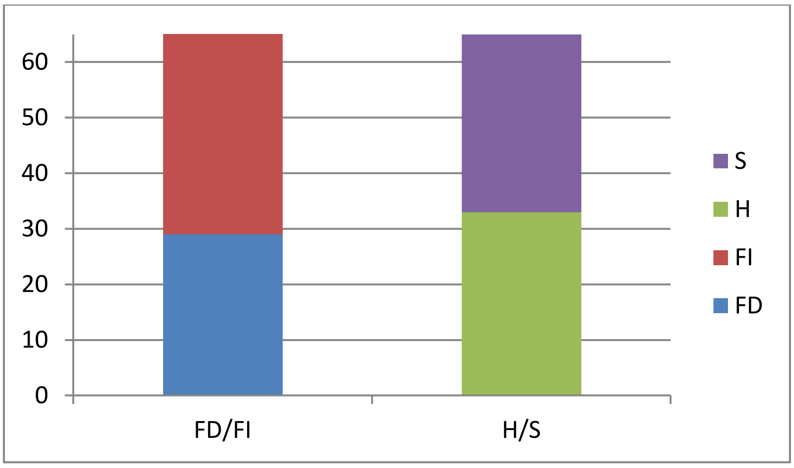



**Figure 3.3. Dataset 2 - Participants' Cognitive Style Distributions**

### 3.3.2. Research Instruments

In addition to the two cognitive styles tests, the CSA and SPQ, described in sections 3.2.2.1 and 3.2.2.2, an exit questionnaire was used to gather data in this study. This is described in the following section.

#### 3.3.2.1. Exit Questionnaire

Subsequently, the participants were then asked to complete a questionnaire that aimed to identify their perceptions and preferences of using the WBI tool. The exit questionnaire consisted of 20 questions (full questionnaire can be found in Appendix B). To fully understand students' preferences, the questionnaire covered a wide range of aspects, including students' comprehension, preferences, and satisfaction or dissatisfaction with the WBI tool, interaction styles, functionality, usability, difficulties and problems. The questionnaire included both positively phrased statements (e.g. '*The map in this tutorial gives a meaningful framework of HTML*') and negatively phrased statements (e.g. '*I found it hard to select relevant information using the map*'). The statements were presented along with a five-point Likert Scale consisting of: 'strongly agree'; 'agree'; 'neutral'; 'disagree'; and 'strongly disagree'. Participants were required to indicate agreement or disagreement with each statement, by placing a check mark at the response that most closely reflected their opinions.

### 3.3.3. Experimental Procedure

After identifying the students' cognitive styles using the CSA and SPQ, the participants were asked to interact with a WBI tool. The tool consisted of an online tutorial, entitled "*How to use HTML*". Although there are many programs that can automatically generate such code, the benefits of using such a tutorial are twofold. Firstly, HTML is at the centre of our online tasks, with it being the building blocks of the Web. As such, the majority of participants are more likely to have had previous exposure to HTML. This not only helps to make the participants more comfortable, but helps to place more of an emphasis on their interaction with the system. Secondly, HTML is one of the simplest and easiest languages to learn. In this way, technical knowledge of a detailed nature was not required as, for example, would be needed if a more complex language was used. Therefore, participant selection was not restricted only to those who were enrolled on technical degree programmes, such as Computer Science or Mathematics.

The tutorial consisted of three main sections and included approximately 80 pages. Among the three sections, Section Two, "*Working with HTML*", contained the majority of content and included 12 sub-topics, each of which was subsequently split into five parts: including an overview, detailed techniques, examples, related skills, and references. The interface was divided into three sections: the title bar, the navigational panel and the main body. The navigation panel was located at the left of the screen and contained several navigational methods, including a main menu, a hierarchical map and an alphabetical index. Figure 3.4 illustrates the interface design of the WBI tool. Through these interface features, students were able to freely navigate within the WBI tool and were provided with multiple options to explore the instructional material in whichever way they chose. Giving students this freedom to choose their preferred method of navigation enabled their preferences for a particular navigation style, level of navigational support and content presentation to be identified. Their perceptions were identified through examining their responses to items in the questionnaire.

## 3.4 Data Analysis

This study aims to analyse the relationships between Field Dependence/ Independence (FD/I) and Holism/Serialism (H/S). The novelty of our data analysis lies within a data mining process, which includes two steps. Firstly, classifiers from two families, Bayesian Networks (BN) and Nearest Neighbour (NN), are used to select relevant questions from the questionnaire for each cognitive style. Secondly, decision trees are created with the selected relevant questions. The decision trees with the highest accuracies are applied to illustrate how FD/I and H/S affect users' preferences for the search engine. Figure 3.4 shows the process of data analysis, with the following sections describing the details.

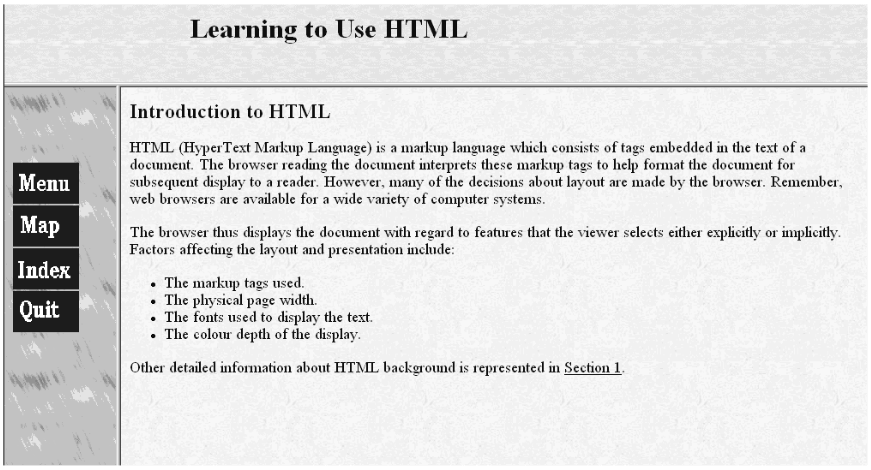



**Figure 3.4: WBI Interface**

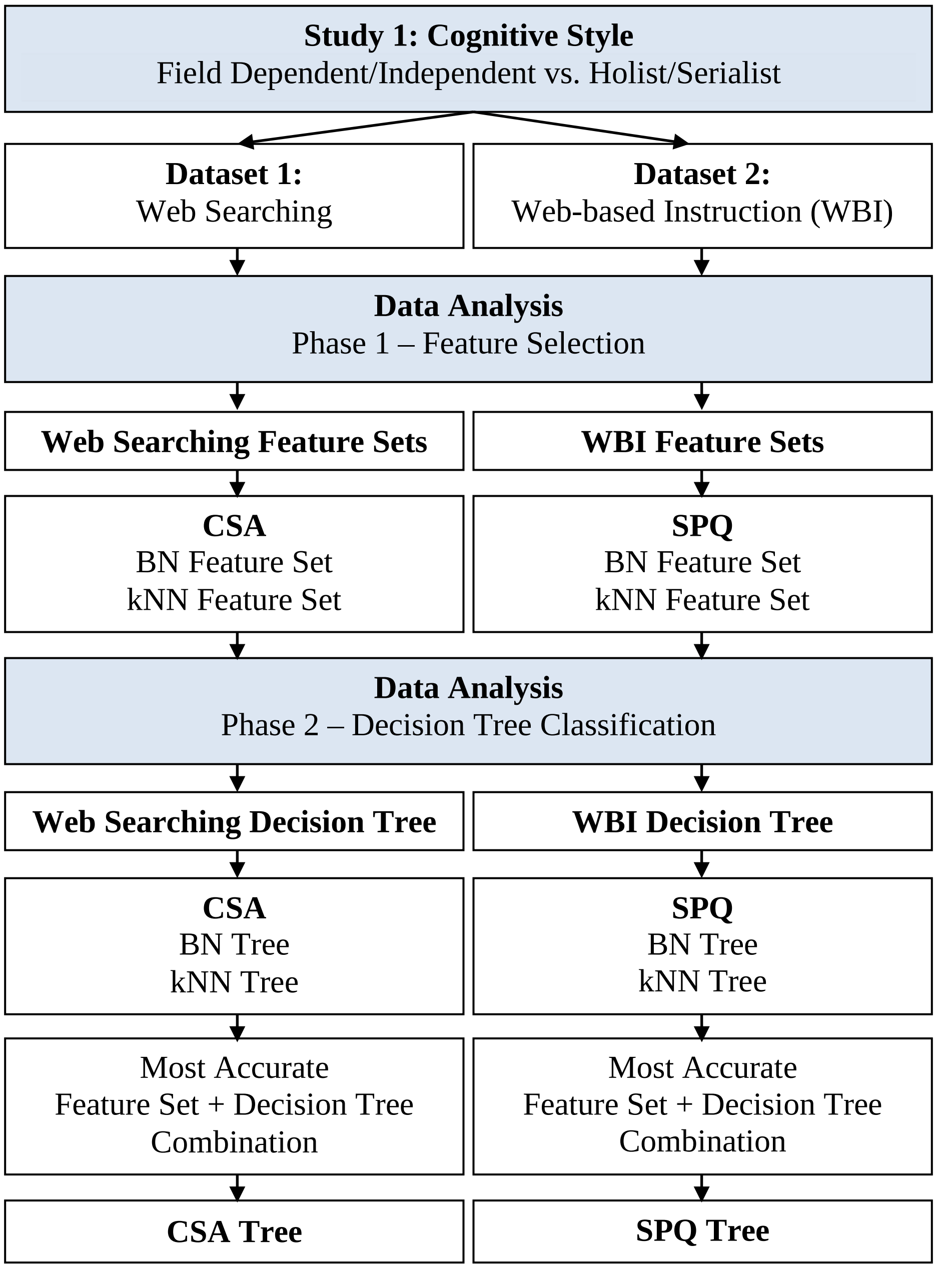



**Figure 3.5. Data Analysis Framework for Study 1**

### 3.4.1 Feature Selection

To select relevant features, classifiers from two different families were used: Bayesian Networks (BN) and k-Nearest Neighbour (KNN). These two families of classifiers were chosen for a number of reasons. Firstly, they are shown to be two of the most influential families of classifiers in a study by Chrysostomou (2008). In addition, these two families were chosen because of the nature of their different biases. The former focuses on features that try to maximise or minimise a scoring metric whereas the latter focuses on selecting features that are deemed ‘closest’ by some sort of imposed distance metric (Gammerman, 1997). Three classifiers from each family were chosen (Table 3.2). Three classifiers from each family were chosen because using three classifiers has been shown to result in highly accurate feature sets (Chrysostomou, 2008). These classifiers were then used to identify highly relevant feature sets, two for each type of application. The four-step method used to attain these four feature sets is described in Figure 3.6.

**Step 1**: Each classifier algorithm was run using 10 fold cross validation. In other words, features were given a score that reflected in how many folds the feature was highlighted as relevant. For example, if the BNC classifier highlighted that Q15 was selected in 8 out of the 10 folds, the score for Q15 from the BNC would be 8.

**Step 2**: For each feature, the three classifier scores in each family were then averaged to give a total score of each feature per family. For example, if Q15 scored 6(BNC), 6(NB) and 3(AODE) in the BN family, the score for Q15 would be 5.

**Step 3**: The average family total was then found by averaging all feature scores collected in Step 2.

**Step 4**: In the final step, any features with a Step 2 score that were higher than the Step 3 score were included in the final feature set. To this end, two feature sets for both types of applications were collected.

**Figure 3.6. Four Step Feature Selection Method**

**Table 3.2: Classifier Families**

| Classifier Family | Classifier |
|---|---|
| Bayesian Network (BN) | Bayesian Networks (BNC)<br>Naïve Bayes (NB)<br>Averaged-One-Dependent Estimates (AODE) |
| k-Nearest Neighbour (KNN) | Nearest Neighbour (NN)<br>k-Nearest Neighbour (KNNC)<br>K-Star (K*) |

### 3.4.2 Decision Tree Classification

Once these four highly relevant feature sets have been collected, the accuracy of them was verified by using decision trees. The classification was conducted using decision trees because they have been used with success in previous research in identifying characteristics of cognitive styles (e.g., Liu and Kešelj 2007). They are also easy to interpret and provide a way of measuring the accuracy of the feature sets (Hsu, Lai and Chiu, 2003). In this way, of the two classifier family feature sets, the two that most accurately represent the preferences of Field Independent/Dependent users and Holists/Serialists, in both of the datasets, was identified.

Among a variety of algorithms that can be used to create decision trees, C4.5 (Quinlan 1993), CART (Breiman et al, 1984) and CN2 (Clark and Niblett, 1989) have been selected. They have been chosen because they are among the most popular, the most established and the best tested in previous research, (e.g. Kim et al, 2002). The analysis with these algorithms consisted of three parts. Firstly, the algorithm that produces the highest average classification accuracy results was identified. Secondly, the feature set with the highest classification accuracy was identified from the algorithm with the highest average. Finally, this tree was used to model the preferences of users and was used to compare with other cognitive style groups.

## 3.5 Findings: Dataset 1

Before beginning the data mining analysis, a simple Pearson's correlation was used to see whether there was an initial relationship between the two cognitive styles. For this dataset, the test revealed that there is indeed a statistically significant ($p=0.00$) relationship between Field Dependent/Independent and Holist/Serialist dimensions. With this in mind, the data mining analysis was run and the results for the Web searching dataset are discussed in the following sections.

### 3.5.1. Feature Selection

As discussed in the previous section, six classifiers from two different families were used to select two different subsets for each cognitive style. The feature sets collected as a result of the feature selection stage are discussed individually by cognitive style and then are compared to each other.

#### 3.5.1.1. Field Dependence/Independence

For the Field Dependent/Independent dimension, the BN and KNN classifiers selected 38 and 46 relevant features (Table 3.3). Both feature sets had 23 features in common:

- Q5 (*"It is easy to understand which icon has been selected."*)
- Q12 (*"The colour scheme matches with your preferences."*)
- Q15 (*"Relevant subject content appears on the same page."*)
- Q16 (*"The results are presented by the levels of the relevance."*)
- Q21 (*"Each window has a title."*)
- Q30 (*"Online instructions appear in a consistent location across screens."*)
- Q41 (*"Bold or larger fonts are used for emphasising important information."*)
- Q43 (*"When you are in trouble, you can receive feedback immediately."*)
- Q44 (*"Messages are brief and unambiguous."*)
- Q52 (*"Sound is used to indicate an error."*)
- Q59 (*"Data inputs are case-blind whenever possible."*)
- Q60 (*"Options used less frequently are located in less-convenient positions."*)
- Q65 (*"A map that shows the structure of the subject content is available."*)
- Q66 (*"The same background colour is used to present content within sections."*)
- Q73 (*"There is a detailed alphabetical index to help locate information."*)
- Q75 (*"The search engine provides you with multiple search options."*)
- Q77 (*"Only information essential to the task is displayed on the screen."*)
- Q78 (*"The title of each page is short, simple, clear, and distinctive."*)
- Q79 (*"Menu items are brief, yet long enough to describe the subject content."*)
- Q80 (*"There are some nice pictures to decorate the layout."*)
- Q82 (*"Online instructions are visually distinctive"*)
- Q88 (*"Help is presented as pop-up window to let you do your task and consult the help system at the same time."*)
- Q90 (*"It is easy to access and return from the help system."*)

#### 3.5.1.2. Holist/Serialist

For the Holist/Serialist dimension, the BN and KNN classifiers selected 35 and 30 features respectively (Table 3.3). The two families selected the following 23 features in common:

- Q9 (*"The names of links and buttons are understandable"*)
- Q10 (*"High contrast colour scheme is applied to present text and background"*)
- Q15 (*"Relevant subject content appears on the same page"*)
- Q21 (*"Each window has a title"*)
- Q25 (*"It is easy to pick up the relevant information directly."*)
- Q26 (*"You can customise the layout, e.g., change the font size or link colour."*)
- Q30 (*"Online instructions appear in a consistent location across screens."*)
- Q31 (*"Icons are clearly labelled."*)
- Q34 (*"Icons and navigation buttons are consistent on all screens."*)
- Q37 (*"When the program presents numbers, integers are right-justified and real numbers are decimal-aligned."*)
- Q39 (*"No more than four colours are utilised in each screen."*)
- Q40 (*"Messages are grammatically correct."*)
- Q43 (*"When you are in trouble, you can receive feedback immediately."*)
- Q44 (*"Messages are brief and unambiguous."*)
- Q55 (*"You will not get stuck when you make a mistake."*)
- Q59 (*"Data inputs are case-blind whenever possible."*)
- Q63 (*"White space creates symmetry, direct attention to appropriate direction"*)
- Q66 (*"The same background colour used to present content within a section."*)
- Q73 (*"There is a detailed alphabetical index to locate specific information."*)
- Q75 (*"The search engine provides you with multiple search options."*)
- Q77 (*"Only information essential to the task is displayed on the screen."*)
- Q78 (*"The title of each page is short, simple, clear, and distinctive."*)
- Q88 (*"Help is presented as pop-up window to let you do your task and consult the help system at the same time."*)

#### 3.5.1.3. Field Dependence/Independence vs. Holist/Serialist

If the findings for Dataset 1 for both types of cognitive style are compared, 11 features are commonly selected by all families (Table 3.4). On close examination, these features fall into three categories: online help, information format and content organisation. This therefore suggests that these are three key aspects of the similarity in the relationship between Field Dependence/Independence and Holism/Serialism.

### 3.5.2. Decision Tree Classification

The second stage of the data mining methodology uses decision tree classification to find the most accurate feature set and then build a tree that is representative of users' preferences for that cognitive style. The following sections discuss the results achieved when carrying out this procedure.

**Table 3.3: Dataset 1: Number of Features Selected per Family**

| Cognitive Style | Feature Sets | | | |
|---|---|---|---|---|
| | **Classifier Family** | **Classifier** | **# of Features Selected** | **# of Features in Final Set** |
| **Field Dependent/ Independent (CSA)** | BN | BNC | 85 | 38 |
| | | NB | 81 | |
| | | AODE | 83 | |
| | KNN | NN | 89 | 46 |
| | | kNN | 86 | |
| | | K* | 88 | |
| **Holist/ Serialist (SPQ)** | BN | BNC | 88 | 35 |
| | | NB | 89 | |
| | | AODE | 80 | |
| | KNN | NN | 74 | 30 |
| | | kNN | 87 | |
| | | K* | 81 | |

**Table 3.4: Commonly Selected Features**

| **Interface Tool** | **Features** |
|---|---|
| Online Help | Q30 (*"Online instructions appear in a consistent location across screens."*) |
| | Q43 (*"When you are in trouble, you can receive feedback immediately."*) |
| | Q44 (*"Messages are brief and unambiguous."*) |
| | Q88 ("*Help is presented as pop-up window to let you do your task and consult the help system at the same time.*") |
| Information Format | Q21 (*"Each window has a title"*) |
| | Q59 (*"Data inputs are case-blind whenever possible."*) |
| | Q66 (*"The same background colour is used to present the content within a section."*) |
| | Q78 (*"The title of each page is short, simple, clear, and distinctive."*) |
| Content Organisation | Q73 (*"There is a detailed alphabetical index to help you locate specific information."*) |
| | Q75 (*"The search engine provides you with multiple search options."*) |
| | Q77 (*"Only information essential to the task is displayed on the screen."*) |

#### 3.5.2.1 Field Dependent/Independent

As can be seen from Table 3.5, the algorithm that performed the best overall was CN2. Although both the BN and KNN feature sets classified with the same accuracy (99.20635%), the BN feature set was chosen as the resulting tree produced was more evenly balanced with similar numbers of branches on both sides of the tree. In addition, this tree highlighted the differences between FI and FD users more clearly.

**Table 3.5. Field Dependent/Independent (CSA) Classification Accuracies**

| Feature Set | Decision Tree Algorithm | | |
|---|---|---|---|
| | *C4.5* | *CART* | *CN2* |
| BN | 93.3333 | 89.1667 | **99.20635** |
| KNN | 96.6667 | 88.3333 | **99.20635** |
| Total Average (%) | 95 | 88.75 | **99.20635** |

The Field Dependent/Independent tree (Figure 3.7) shows that there are six features that were key considerations in identifying the differences in preference of FI and FD users. In addition, the decision rules derived from this tree are shown in Table 3.6. The root node in the tree is Q75 (“*search engine provides multiple options*”), which very clearly categorises the two types of users. Users who thought having multiple options for the search engine was ‘very unimportant’ or ‘unimportant’ are FI, while users who thought it ‘important’ or ‘very important’ are FD. Q59 (“*data inputs are case-blind whenever possible.*”) also neatly splits the preferences of FI and FD users, with the latter preferring to have inputs case-blind over the former. Q15 (“*relevant subject content appears on the same page*”) and Q16 (“*the results are presented by the levels of the relevance*”) also illustrate the differences in preferences between FD and FI users. Users who preferred results to be presented by level of relevance and for relevant content to have appeared on the same page were FI users, whereas FD users thought that having results presented by level of relevance was ‘unimportant’. Although Q36 (“*messages are appropriate, inoffensive, non-hostile or violent*”) and Q65 (“*a map that shows the structure of the subject content is available*”) did not clearly show the split between preferences for the two different types of users, nevertheless they showed that FI users thought it ‘very important’ that messages

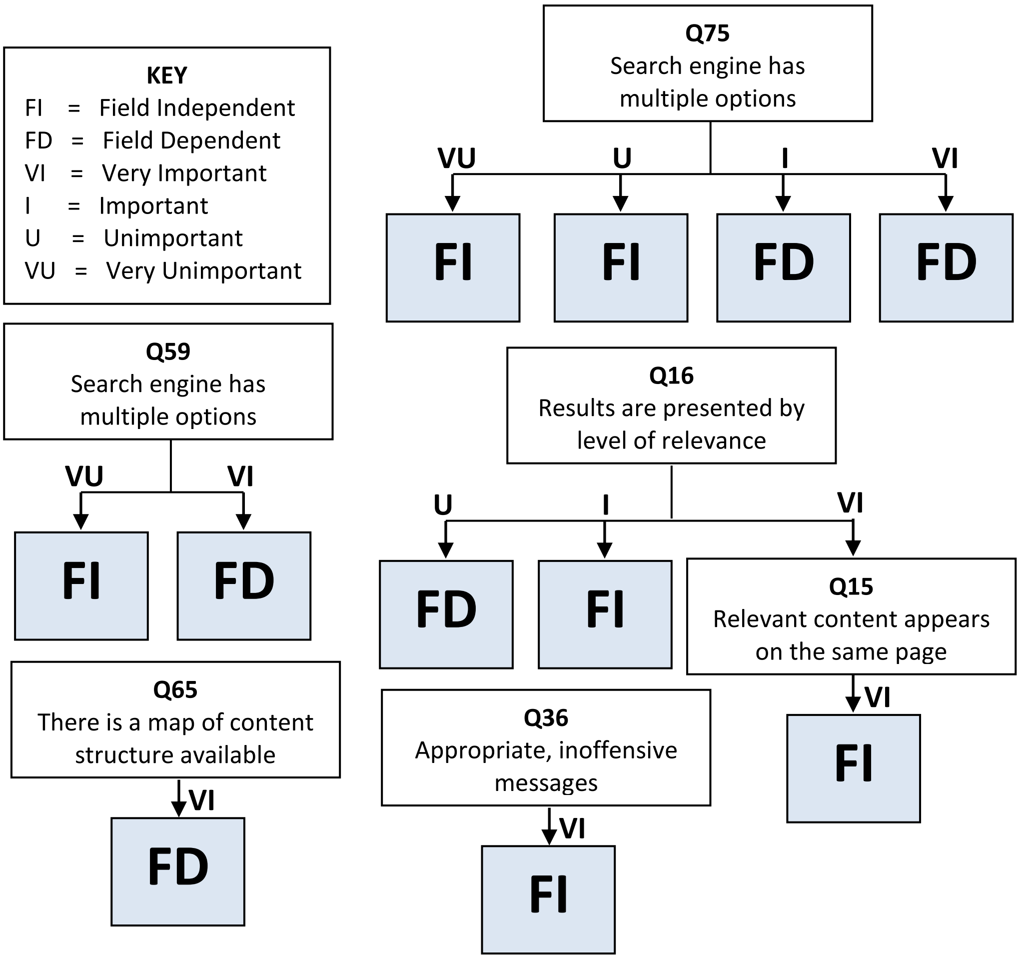



**Figure 3.7: Dataset 1 – Field Dependent/Independent Decision Tree**

**Table 3.6. Decision Rules for Field Dependent/Independent Decision Tree**

| Dataset | FI/FD | Decision Rules |
|---|---|---|
| Dataset 1: Web Searching | FD | If user thinks it very important or important for the search engine to have multiple options, user is Field Dependent. |
| | FD | If user thinks that it is unimportant to have results presented by relevance, they are Field Dependent. |
| | FD | If user thinks it is very important to have case-blind data inputs, they are Field Dependent. |
| | FD | If user thinks it is very important to have a map of the content structure, they are Field Dependent. |
| | FI | If user thinks it is unimportant or very unimportant for the search engine to have multiple options, they are Field Independent. |
| | FI | If the user thinks it is important to have results presented by level of relevance, they are Field Independent. |
| | FI | If the user thinks it is very important to have results presented by level of relevance and it is very important for relevant content to appear on the same page, they are Field Independent. |
| | FI | If the user thinks it is very unimportant to have case-blind data inputs, they are Field Independent. |
| | FI | If the user thinks it is very important to have appropriate, inoffensive messages, they are Field Independent. |

were appropriate and inoffensive and FD users thought it was ‘very important’ to have a map of the content structure available. Interestingly, four of the five features, i.e., Q15, Q16, Q65 and Q75, identified in this tree were in the small selection of features commonly selected by all of the classifiers in both families for FD/I in Step 1. This suggests that these four features were the most relevant features that classify FD and FI users the most accurately.

#### 3.5.2.2 Holist/Serialist

The CART algorithm produced the highest average classification accuracy of 87.08335 (Table 3.7), so the most accurate feature set from this algorithm, BN with 87.08335%, was then used to build a decision tree. As shown in Figure 3.8, and the decision rules displayed in Table 3.8, nine features played a key role in determining the differences between Holists and Serialists. Among these nine features, four, i.e., Q9, Q21, Q37 and Q75, were within the set of seven features commonly selected as relevant by all of the classifiers in Step 1. This suggests that these four features were the most relevant features that classify Holists and Serialists. In particular, Q75 is very prominent and can be seen at the top level, signifying that this was the most important feature that can determine the differences between the two types of user most accurately.

However, the differences between Holists and Serialists cannot be defined through the use of just one question, like the FI/FD tree showed. Conversely, in the Holist/Serialist tree a combination of questions was needed to discover the differences between these two types of user. For example, if a user thought it was ‘very unimportant’ or ‘important’ for search engines to have multiple options (Q75), and they thought that it was important that numbers were right hand justified and decimal aligned (Q37), and it was very important for pages to be clearly marked with a label that indicated the subject content (Q1), the user was likely to be a Holist. Conversely, if a user thought Q75 was ‘unimportant’ or ‘very important’ and they thought it ‘important’ that relevant subject content appeared on the same page (Q15), then the user was likely to be a Serialist.

**Table 3.7: Holist/Serialist (SPQ) Classification Accuracies**

| Feature Set | Decision Tree Algorithm | | |
|---|---|---|---|
| | *C4.5* | ***CART*** | *CN2* |
| BN | 85 | **87.5** | 75.83333 |
| KNN | 85 | **86.6667** | 76.6667 |
| Total Average | 85 | **87.08335** | 76.25002 |

**KEY**
FI = Field Independent
FD = Field Dependent
VI = Very Important
I = Important
U = Unimportant
VU = Very Unimportant

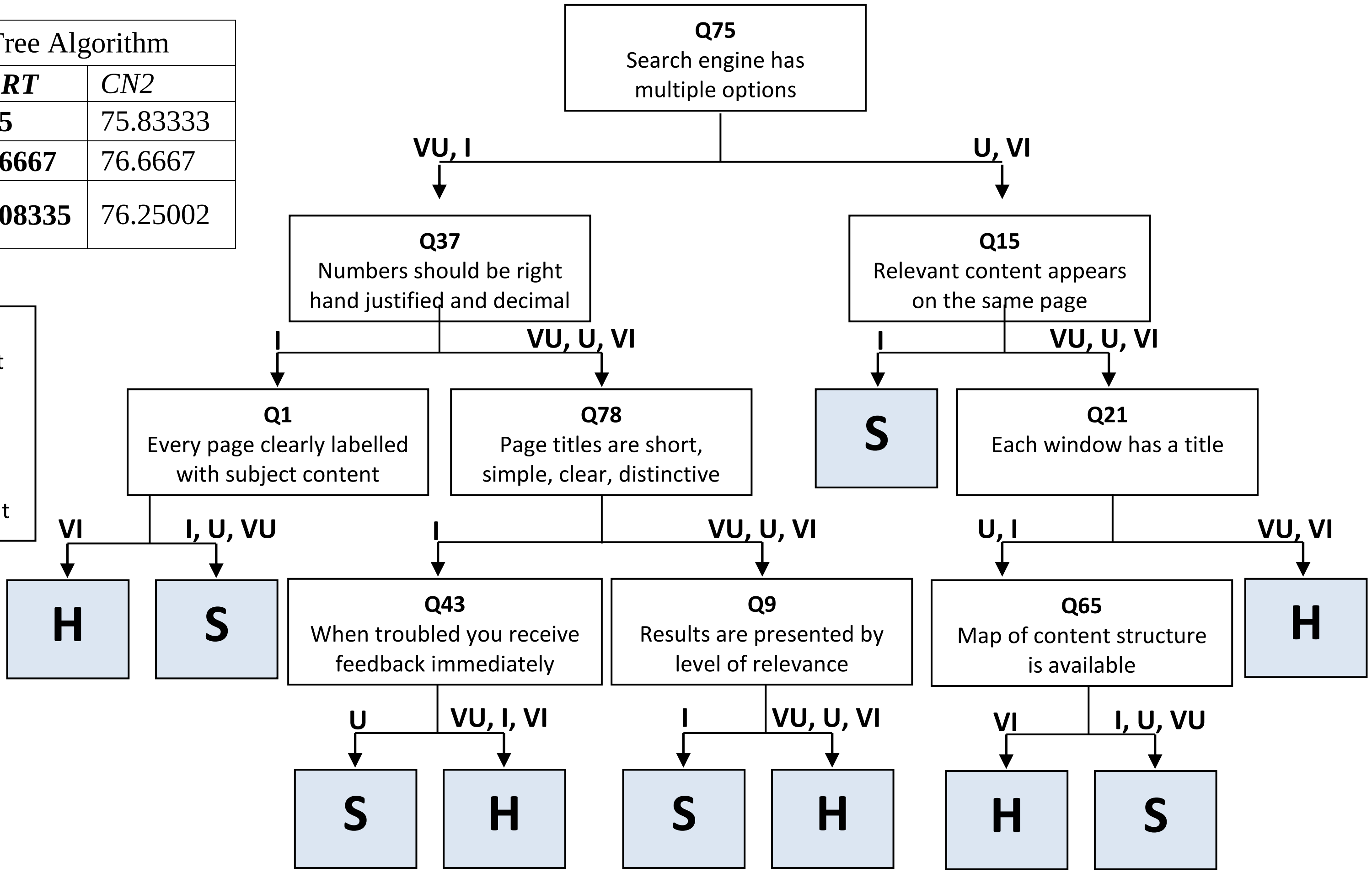



**Figure 3.8: Dataset 1 Holist/Serialist Decision Tree**

**Table 3.8. Decision Rules for Holist/Serialist Decision Tree**

| Dataset | H/S | Decision Rules |
|---|---|---|
| Dataset 1: Web Searching | H | If the user thinks it very unimportant or important for search engine to have multiple options, AND it is important that numbers should be right hand justified and decimal aligned, AND it is very important that page is clearly labelled with subject content, they are Holist. |
| | S | If the user thinks it is unimportant or very important for the search engine to have multiple options, AND it is important for relevant content to appear on the same page, they are Serialist. |
| | H | If the user thinks it unimportant or very important for search engine to have multiple options, AND it is very unimportant, unimportant or very important for relevant content to appear on the same page, AND it is very unimportant or very important for each window to have a title, they are Holist. |
| | H | If the user thinks it unimportant or very important for the search engine to have multiple options, AND it is very unimportant, unimportant or very important for relevant content to appear on the same page, AND it is unimportant or important for each window to have a title, AND it is very important to have a map of the content structure, they are Holist. |
| | S | If the user thinks it is unimportant or very important for the search engine to have multiple options, AND it is very unimportant, unimportant or very important for relevant content to appear on the same page, AND it is unimportant or important for each window to have a title, AND it is very unimportant, unimportant or important to have a map of the content structure, they are Serialist. |
| | S | If the user thinks it is very unimportant or important for the search engine to have multiple options, AND it is important that numbers should be right hand justified and decimal aligned, AND it is very unimportant, unimportant or important that every page is clearly labelled with subject content, they are Serialist. |
| | S | If the user thinks it is very unimportant or important for the search engine to have multiple options, AND it is very unimportant, unimportant or very important that numbers should be right hand justified and decimal aligned, AND it is important that the title of each page is short, simple, clear and distinctive, AND it is unimportant that feedback is received immediately, they are Serialist. |
| | H | If the user thinks it is very unimportant or important for search engine to have multiple options, AND it is very unimportant, unimportant or very important that numbers should be right hand justified decimal aligned, AND it is important that the title of page is short, simple, clear and distinctive, AND it is very unimportant, important/very important that feedback received immediately, they are Holist. |
| | S | If the user thinks it is very unimportant or important for the search engine to have multiple options, AND it is very unimportant, unimportant or very important that numbers should be right hand justified and decimal aligned, AND it is very unimportant, unimportant or very important that the title of each page is short, simple, clear and distinctive, AND it is important that the names of links and buttons are understandable, they are Serialist. |
| | H | If the user thinks it is very unimportant or important for the search engine to have multiple options, AND it is very unimportant, unimportant or very important that numbers should be right hand justified and decimal aligned, AND it is very unimportant, unimportant or very important that the title of each page is short, simple, clear and distinctive, AND it is very unimportant, unimportant or very important that the names of links and buttons are understandable, they are Holist. |

In summary, it seems that the differences between Holists and Serialists are not as clearly divided as those of FI and FD users. More specifically, the results observed from the Holist/Serialist tree seem at times controversial as some of the branches do not follow as logical an argument as the FI/FD tree. For example, the branch of Q75 groups users together whose preferences are ‘very unimportant’ and ‘important’, or ‘unimportant’ and ‘very important’. This is probably related to the instrument used to identify the Holists and Serialists in this study. As described in Section 3.2.2., the SPQ was used to identify the participants as Holists and Serialists, but the reliability of the SPQ is not yet validated, which may thus influence the accurateness of the Holist/Serialist tree and the logicalness of each condition.

#### 3.5.2.3 Field Dependent/Independent vs. Holist/Serialist

If the two decision trees (Figure 3.7 and 3.8) are compared, it is clear to see that there are two features that appear in both trees: Q9 and Q75. The most prominent similarity that can be seen between the two cognitive styles is Q75. It is quite clear that FI and FD users were divided through their preferences for this question. Although the Holism/Serialism classification cannot be as easily determined as FD/I, nevertheless Q75 was used as the root node – the one feature that most accurately separated users into Holists and Serialists. This would suggest that the preference for having multiple options in the search engine (Q75) was very relevant to both cognitive styles. More specifically, FD users thought it was important to have multiple options while FI users thought it was unimportant. In other words, FD users prefer to have lots of options available to them as they prefer to take a global approach (Witkin, *et al*., 1977). In comparison, FI users do not think that having many options available is important because they tend to be more analytical and prefer to focus only on a specific part. Likewise, Holists prefer to have more options than Serialists as the former tend to take a holistic approach, whilst the latter focus more on procedural details (Pask, 1979). To address such differences, there is a need to provide personalised interfaces. The personalised interface for Field Dependent users or Holists would display all of the available options. On the other hand, only the options that are relevant to their current tasks would be shown to Field Independent users or Serialists. In other words, any irrelevant options will be hidden so that the Field Independent users are not confused.

Two other similarities can be seen in the two trees. Firstly, Q65 appears in both trees. With the FD/I tree, users who thought it was very important to have a map that showed the structure of the content were FD users. With the results from Holism/Serialism, users who thought this important were Holists. In other words, there is a similarity between FD users and Holists. Secondly, Q15 also appears in both trees. In the FD/I tree, users who thought it 'very important' that relevant subject content appeared on the same page were FI users. However, in the Holism/Serialism tree, users who thought this 'very important' could not be as easily classified as Holist or Serialist without further looking at either Q21 and/or Q65. It reveals that the similarity between FI users and Serialists is not as clear as FD users and Holists.

### 3.5.3 Statistical Significance

In addition to the aforesaid data mining methods, other conventional methods were also used to identify the relationship between FD/I and Holism/Serialism. Table 3.9 presents the number of the participants in each dimension of the cognitive style. As shown in this table, there are more FD/Holists than FD/Serialists and there are more FI/Serialists than FI/Holists. This demonstrates that there is indeed a marked similarity between FD users and Holists and between FI users and Serialists.

**Table 3.9. Distribution of Participants' Cognitive Styles**

| | Field Independent | Field Dependent | Total |
|---|---|---|---|
| **Holist** | 25 | 38 | 63 |
| **Serialist** | 34 | 23 | 57 |
| **Total** | 59 | 61 | 120 |

Furthermore, statistical tests were run on the original dataset including all 90 features because they have been widely used in the studies of cognitive styles (e.g., Ford, *et al.*, 2005). Thus, we check whether the features identified in Section 4.2 were also statistically significant. More specifically, these features were checked against the results of the t-test (Table 3.10). Although there were more features identified as relevant to Holists and Serialists, the results of the t-test show that only three of the nine are statistically significant ($p<0.05$). However, five of the six features identified as relevant to Field Dependent/Independent users are shown to be statistically significant. This suggests the possibility of the results related to Holism/Serialism not being so reliable. One possible reason for this could be that FD/I was identified by using the CSA whilst Holism/Serialism was categorised by the SPQ. The former is more established and better tested than the latter. Alternatively, the preferences of FD and FI users seem to be clearer and more easily identifiable than those of Holists/Serialists. However, Q75 is statistically significant to both cognitive styles.

**Table 3.10. Statistical Significance of Features ($p<0.05$)**

| Field Dependent/Independent | Holism/Serialism |
|---|---|
| Q15<br>Q16<br>Q59<br>Q65<br>Q75 | Q9<br>Q37<br>Q75 |

### 3.5.4 Summary

In summary, the results from Dataset 1 identify three major findings. Firstly, the results confirm that there is indeed a fundamental link between the two cognitive styles. This is demonstrated by the repeated emphasis of Q75 in all steps. In Step 1 (feature selection), this feature was the only commonality in the feature sets selected for both cognitive styles. Step 2 (classification) showed that Q75 clearly separates the preferences of FI and FD users and appeared as the root node in the Holist/Serialist tree, allowing for the most accurate separation of Holist and Serialist users' preferences. In addition, further analysis finds this question statistically significant to both cognitive styles.

Secondly, we have confirmed that there are some similarities between FD users and Holists and between FI users and Serialists. However, the results suggest that the similarities between the latter are more prominent than the former. For example, the classification step highlights the similarity between FD users and Holists through the appearance of Q75 and Q65 in both of the decision trees. Nonetheless, the similarity between FI users and Serialists only lies within Q75. Finally, the third and perhaps most significant finding clearly relates the characteristics of users to preferences regarding search engines. More specifically, the results show that the preferences FI and FD users have for the interface design of search engines can be separated more clearly and logically than those of Holists and Serialists, suggesting that identifying the different preferences of Holists and Serialists is more complex. For example, the classification step shows that Q75 and Q59 allow for a clear split in preferences between FI and FD users, with the former finding them unimportant and the latter users finding them important. However, to identify the preferences of Holists or Serialists, a combination of more than one feature is needed.

## 3.6 Findings: Dataset 2 – Web-Based Instruction Tool

### 3.6.2 Feature Selection

As discussed in the previous section, six classifiers were used from two different families to select two relevant subsets for each cognitive style.

#### 3.6.2.1 Field Dependent/Independent

For the Field Dependent/Independent dimension, the BN and KNN classifiers selected 9 and 8 relevant features respectively (Table 3.11). Both feature sets had five features in common:

- Q9 (*'It is hard to use back/forward buttons'*),
- Q11 *('the links provided in this tutorial help me to discover relationships between different topics')*,
- Q14 ('*After using this system I can easily use my knowledge to design home pages*'),
- Q15 ('*I found it hard to select relevant information using the map'*),
- Q19 ('*This tutorial can be used sufficiently well without any instructions'*).

#### 3.6.2.2 Holist/Serialist

The BN and NB classifiers selected 6 and 10 features relevant to Holist/Serialist (Table 3.11). Again, both of these feature sets had five features in common:

- Q2 ('*Examples given in this tutorial are not practical*'),
- Q7 ('*I would like to have had more examples*'),
- Q9 ('*It is hard to use back/forward buttons*'),
- Q15 ('*I found it hard to select relevant information using the map*'),
- Q18 ('*It is easy to find specific information for a task with the index*').

**Table 3.11: Dataset 2 – Number of Features Selected per Family**

| Cognitive Style | Feature Sets | | | |
|---|---|---|---|---|
| | **Classifier Family** | **Classifier** | **# of Features Selected** | **# of Features in Final Set** |
| **Field Dependent/ Independent (CSA)** | BN | BNC | 17 | 9 |
| | | NB | 15 | |
| | | AODE | 10 | |
| | kNN | NN | 16 | 8 |
| | | KNN | 20 | |
| | | K* | 14 | |
| **Holist/ Serialist (SPQ)** | BN | BNC | 13 | 6 |
| | | NB | 10 | |
| | | AODE | 14 | |
| | kNN | NN | 16 | 10 |
| | | KNN | 10 | |
| | | K* | 13 | |

#### 3.6.2.3 Field Dependent/Independent vs. Holist/Serialist

If these findings are compared, two features appear as both relevant to Field Dependent/Independent and Holist/Serialist: Q9 and Q15. This suggests that these two features are the most relevant to each of the cognitive style dimensions. It is worth noting that both of these features refer to the way a user prefers to navigate through the subject. It implies that these two dimensions of cognitive styles have a close relationship with users' navigation preferences.

### 3.6.3 Decision Tree Classification

Once the feature sets were identified, classification was then performed to identify the most accurate feature set. This feature set was then used to build a decision tree, which can illustrate users' preferences of WBI programs. Table 3.12 shows the classification accuracies for the Field Dependent/Independent feature sets. The C4.5 algorithm had the best overall average and performed the most accurately using the KNN feature set. Table 3.14 shows the classification accuracies for the Holist/Serialist feature sets. The CN2 algorithm performed the most accurately, with both feature sets having the same highest classification accuracies.

**Table 3.12: Field Dependent/Independent (CSA) Classification Accuracies**

| Feature Set | Decision Tree Algorithm | | |
|---|---|---|---|
| | ***C4.5*** | *CART* | *CN2* |
| BN | **87.6923** | 81.3218 | 90.76923 |
| KNN | **95.3846** | 81.3218 | 86.15385 |
| Total Average | **91.53845** | 81.3218 | 88.46154 |

Having identified the feature sets that produced the highest classification accuracies, they are assumed to be the ones that most accurately contain the characteristics of each cognitive style group. Therefore, these feature sets are presumed to most accurately represent the different types of users' preferences for WBI. Figure 3.9 shows the decision tree produced for the most accurate Field Dependent/Independent feature set with Table 3.13 showing the decision rules derived from this tree. Figures 3.10 and 3.11 show the decision trees produced from the two equally accurate Holist/Serialist feature sets, with Table 3.15 showing the decision rules derived from these trees. When these chosen decision trees are compared, three main similarities can be seen through features Q9, Q6 and Q18. Q9 ('*It is hard to use back/forward buttons*') appears prominently in both the FI/FD and the Holist/Serialist trees. In both, the differences between the two categories are clear. For example, in the FI/FD tree, FI users are shown to strongly disagree with Q9 whereas FD users agree or strongly agree. Alternatively, in the Holist/Serialist tree, Holists are shown to agree whereas Serialists are shown to disagree/strongly disagree. Back/forward buttons are used to help users explore the content in a non- linear way. This finding suggests that the Field Independent users and Serialists in this study show a similar preference for linear navigation, whereas Field Dependent users and Holists prefer to navigate in a much more non-linear manner. Such a finding agrees with the known characteristics of Holists and Serialists. Holists have been shown to prefer a much more global approach, following links to get an overall understanding before going into detail, whereas Serialists prefer linear navigational methods as they tend to take a local analytical approach that covers each area in detail before moving on to the next (Pask, 1976).

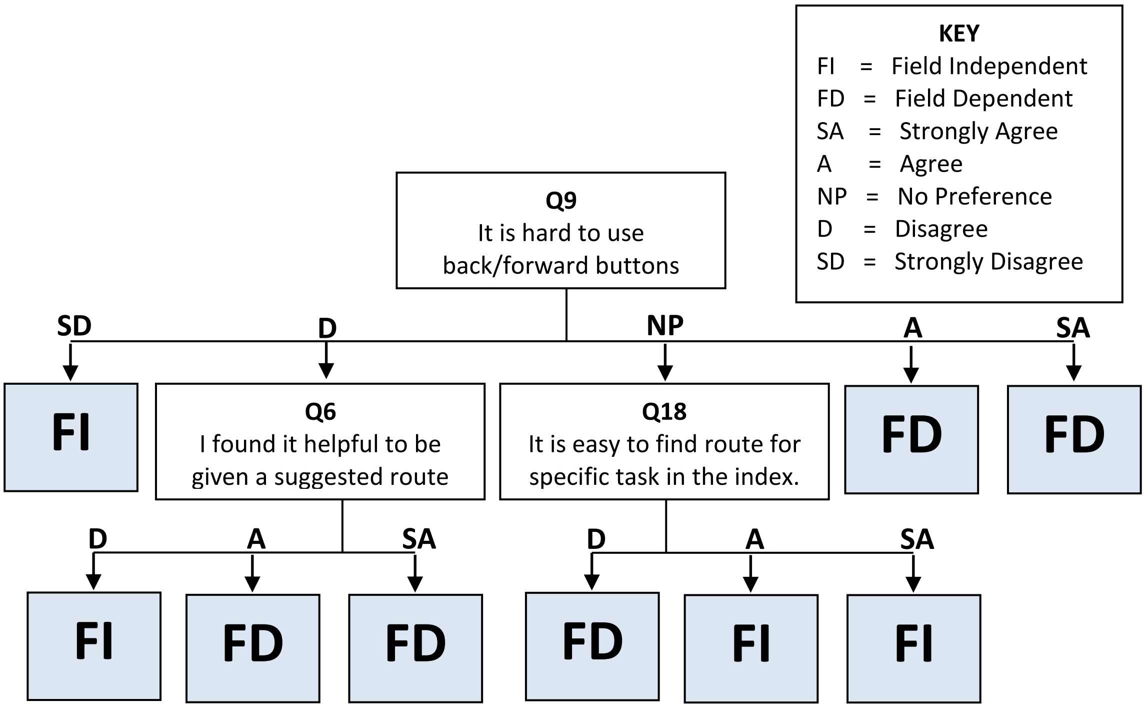



**Figure 3.9. Dataset 2 - Field Dependent/Independent Decision Tree**

**Table 3.13. Decision Rules for Field Dependent/Independent Tree**

| Dataset | FI/FD | Decision Rules |
|---|---|---|
| Field Dependent / Independent | FI | If the user strongly disagrees that it is hard to use backwards and forwards buttons, they are Field Independent. |
| | FD | If the user agrees or strongly agrees that it is hard to use backwards and forwards buttons, they are Field Dependent. |
| | FI | If the user disagrees that it is hard to use backwards and forwards buttons, AND disagreed that it would have been more helpful if they were given a suggested route through the tutorial, they are Field Independent. |
| | FD | If the user disagrees that it is hard to use backwards and forwards buttons, AND agreed or strongly agreed that it would have been more helpful if they were given a suggested route through the tutorial, they are Field Dependent. |
| | FD | If the user has no preference on the difficulty on using backwards and forwards buttons, AND disagrees that it is easy to find a route for a specific task within the index, they are Field Dependent. |
| | FI | If the user has no preference on the difficulty on using backwards and forwards buttons, AND agreed or strongly agrees that it that it is easy to find a route for a specific task within the index, they are Field Independent. |

Q6 ('*I would have found it more helpful to be given a suggested route through this tutorial*') also appears in both trees. The FI/FD tree shows that FI users disagree with this question whereas FD users agree or strongly agree. The Holist/Serialist tree shows that Holists strongly disagree with this question and Serialists strongly agree. Unlike the results of Q9, this indicates that FI users and Holists have similar preferences whereas FD users are like Serialists. More specifically, the former can explore the content on their own while the latter need more guidance in finding their way around a topic.

Q18 ('*It is easy to find specific information for a task with the index*') is another feature selected by both trees. Both FI users and Serialists strongly agree or agree with this statement, whilst FD users and Holists disagree. The index used in this study lists all of topics in an alphabetical order so that learners can easily locate specific information. This suggests that such a mechanism is useful to FI users and Serialists but it may not be suitable for FD users and Holists. This may be due to the fact that FD users and Holists are interested in a global picture of the content, instead of a specific item.

In addition to these similarities, there are three other features that appear on the Holist/Serialist tree. Q11 ('*the links provided in this tutorial help me to discover relationships between different topics*') is selected in the Holist/Serialist tree, with Holists strongly disagreeing to this statement and Serialists strongly agreeing to it. It is in the nature of Holists to jump to different topics (Pask, 1979) so links are helpful for them. On the other hand, Serialists tend to study topics sequentially so there may be no need for them to use links. Although this feature does not appear in the chosen Field Dependent/Independent tree, it is shown in the tree that was drawn using CN2 and the KNN family feature set. It can be seen in this tree that FD users agree with Q11. This suggests that FD users may have a similar preference to Holists for this question. Q2 ('*Examples given in this tutorial are not practical') and Q7 ('I would like to have had more examples*') were also selected in the Holist/Serialist tree but not in the Field Dependent/Independent tree. Examples are another way of presenting content so this suggests that the content presentation may be more relevant to Holists and Serialists than to Field Dependent and Field Independent users.

**Table 3.14: Holist/Serialist (SPQ) Classification Accuracies**

| Feature Set | Decision Tree Algorithm | | |
|---|---|---|---|
| | *C4.5* | *CART* | ***CN2*** |
| BN | 72.3077 | 79.79925 | **80** |
| KNN | 73.8462 | 76.79925 | **80** |
| Total Average | 73.07695 | 78.29925 | **80** |

**KEY**
H = Holist
S = Serialist
SA = Strongly Agree
A = Agree
NP = No Preference
D = Disagree
SD = Strongly Disagree

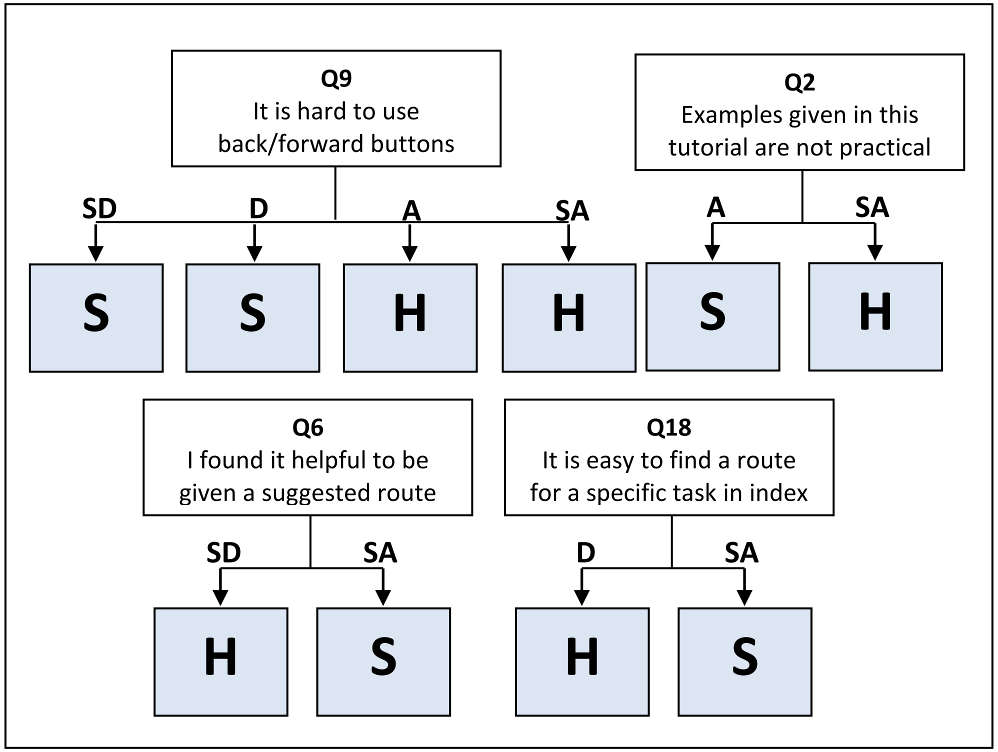



**Figure 3.10. Dataset 2 - Holist/Serialist Decision Tree 1**

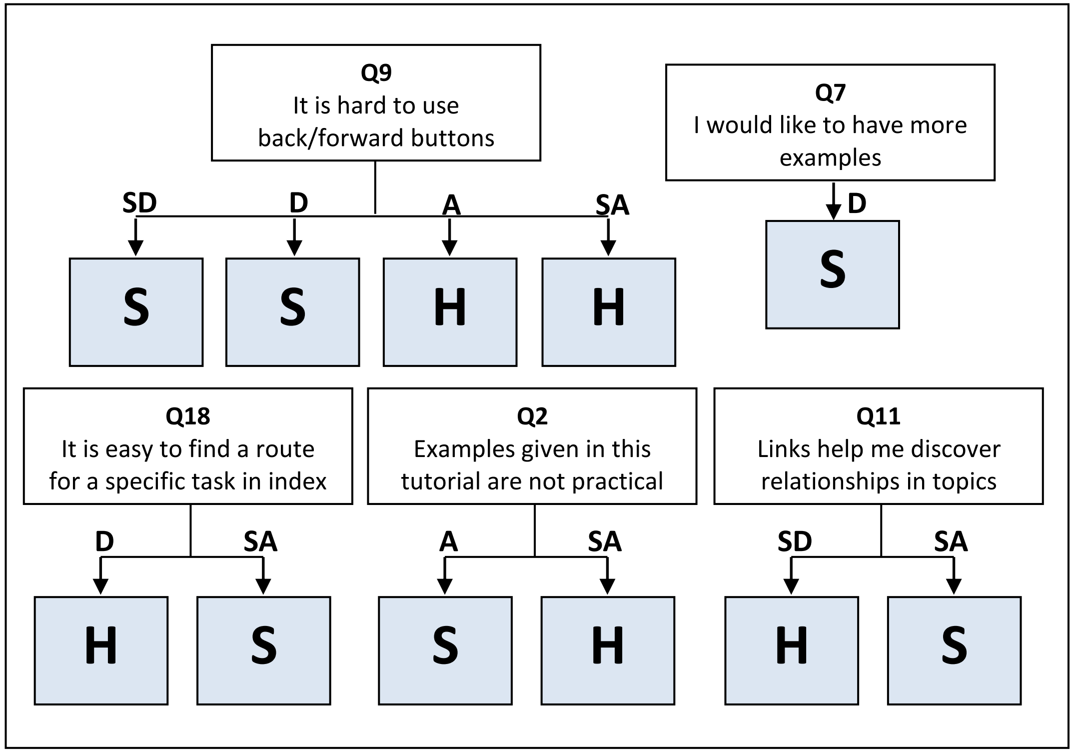



**Figure 3.11. Dataset 2 - Holist/Serialist Decision Tree 2**

**Table 3.15. Decision Rule Table for Holist/Serialist Trees**

| Dataset | H/S | Decision Rules |
|---|---|---|
| Holist/ Serialist | H | If the user agrees or has no preference that it is hard to use backwards and forwards buttons, they are Holist. |
| | H | If the user strongly disagrees that the links provided in this tutorial helped them to discover the relationships between the different topics, they are Holist. |
| | H | If the user strongly agrees that the examples given in this tutorial are not practical, they are Serialist. |
| | H | If the user strongly disagrees that it is easy to find a route for a specific task within the index, they are Holist. |
| | S | If the user strongly disagrees or disagrees that it is hard to use backwards and forwards buttons, they are Serialist. |
| | S | If the user disagrees that they would have liked more examples, they are Serialist. |
| | S | If the user strongly agrees that the links provided in this tutorial helped them to discover the relationships between the different topics, they are Serialist. |
| | S | If the user agrees that the examples given in this tutorial are not practical, they are Serialist. |
| | S | If the user strongly agrees that it is easy to find a route for a specific task within the index, they are Serialist. |

### 3.6.4 Statistical Significance

As with the other dataset, additional statistical methods were used to verify the relationship between the two cognitive styles for Dataset 2. Table 3.16 shows the number of participants belonging to each dimension of cognitive style. As with the other dataset, there is a higher proportion of Field Dependent users that are Holists and a higher proportion of Field Independent users that are also Serialists. Again, this demonstrates that there is more of a significant similarity between FD users and Holists and between FI users and Serialists.

**Table 3.16. Distribution of Participants' Cognitive Styles**

| | Field Dependent | Field Independent | Total |
|---|---|---|---|
| **Holist** | 20 | 16 | 36 |
| **Serialist** | 13 | 16 | 29 |
| **Total** | 33 | 32 | 65 |

In addition, an ANOVA test was run on the original dataset that included all of the 20 features. The results show that all of the features in this dataset were found to be statistically significant ($p<0.05$) for the Field Dependent/Independent dimension, whereas only one of the twenty was found to be statistically significant for the Holist/Serialist dimension (Table 3.17). Such a finding suggests a number of things. Firstly, it is well known that the CSA is a more established method of deciding whether a user is Field Dependent or Field Independent, whereas the SPQ is less tried and tested for allocating users as Holist or Serialist in previous research. It could be that the SPQ is a less reliable method and therefore attains less reliable results. Secondly, it could be that the questions were not able to correctly capture the preferences of Holist and Serialist users. Future research should perhaps look at investigating this point further.

**Table 3.17: Statistical Significance of Dataset 2**

| PREF Dataset (Web-based Learning) | |
|---|---|
| **CSA** | **SPQ** |
| Q1, Q2, Q3, Q4,Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18, Q19, Q20 | Q11 |
| T = 20 | T = 1 |

### 3.6.5 Summary

In summary, the findings for Dataset 2 agree with those of Dataset 1. In general, Field Independent individuals and Serialists have similar preferences, as do Field Dependent individuals and Holists. As with Dataset 1, one feature (Q9) is highlighted in all four steps and is identified as a key feature that allows for the most accurate separation between Field Dependent/Field Independent users and between Holists/Serialists. However, there were contradictions found to this rule, specifically in Q6 (‘*I would have found it more helpful to be given a suggested route through this tutorial*’). Field Independent users and Holists disagree with this question whilst Field Dependent users and Serialists agree.

## 3.7 Conclusion

This chapter presented a study that aimed to investigate the influence of cognitive style on users' preferences for Web-based applications. More specifically, the relationship between Field Dependent/Independent and Holist/Serialist was analysed with the aim of identifying whether previous researchers are correct in their assumptions that Field Dependent individuals have similar preferences to Holists, and Field Independent individuals have similar preferences to Serialists. As shown in Figure 3.12, the findings from both datasets show that in general this is true, although some contradictions can be noted. However, the relationships between Field Dependent users and Holists and between Field Independent users and Serialists is stronger than the relationships between Field Independent users and Holists and Field Dependent users and Serialists. Consequently, previous literature identifying the differences of Field Dependent and Field Independent users can perhaps be used to suggest the differences between Holists and Serialists, but bearing in mind the contradictions to the relationship.

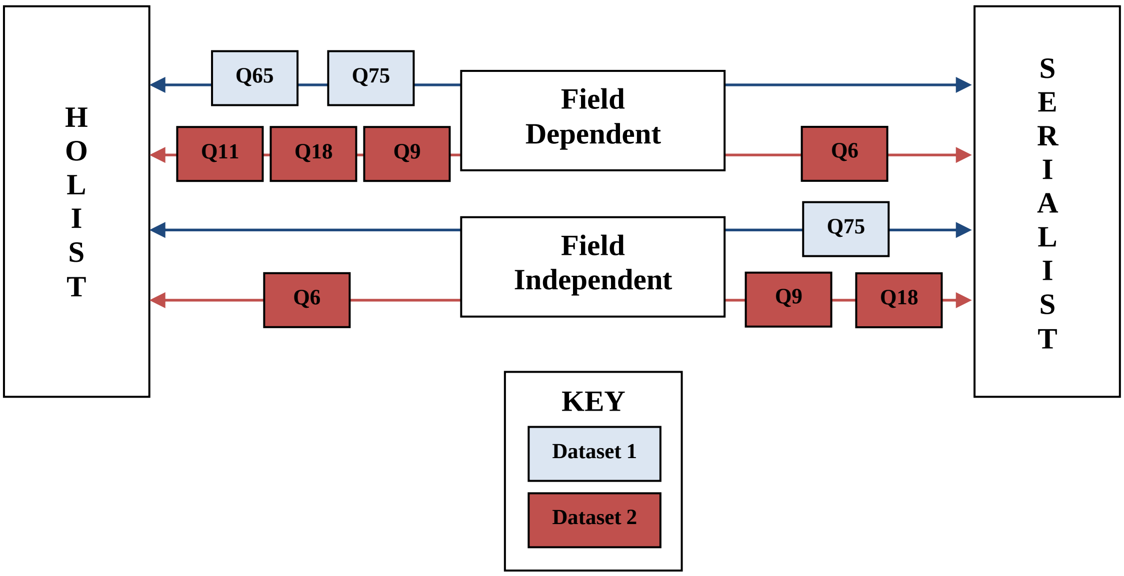



**Figure 3.0-12. Relationship between Field Dependence/Independence and Holism/Serialism**

# Chapter 4 – Study 2: Internet Experience vs. Computer Experience

## 4.1. Introduction

This chapter will describe the second of the three studies examined within this thesis. This study focussed on the effect that previous system experience had on users' preferences of Web-based applications. More specifically, users' levels of previous Internet Experience and Computer Experience were examined. As stated in Chapter 2, although these are difference dimensions of system experience, researchers often group these two types of experience together, with the assumption that Computer experts will act the same as Internet Experts and Computer novices will act the same as Internet novices. Therefore, the goal of this study is to examine the relationship between the two dimensions to see if their similarities are such that they can be grouped as one. Thus, we propose the following research question as our goal for this particular study:

*To what extent do computer experts have the same preferences as Internet experts and computer novices have the same preferences as Internet novices?*

An investigation of this kind will make contributions in several areas. Firstly, it will provide further understanding of the characteristics of experts and novices in both Internet and Computer Experience. Furthermore, we can identify whether there is a relationship between the two types of experience. In most Web-based applications, the difference between these two types of experience are not distinguished, with different interfaces provided for experts and novices only, rather than Internet experts, Computer experts, Internet novices and Computer novices. If there is indeed a similar relationship between the two, then this type of grouping is acceptable. However, if the relationship is not similar, further research needs to commence so that designers can provide effective interfaces to differentiate between the different types of user. Figure 4.1 presents the framework of this study. Using the same integrated data analysis technique that is described in the previous chapter, the following sections go on to discuss the findings and conclusions of this study.

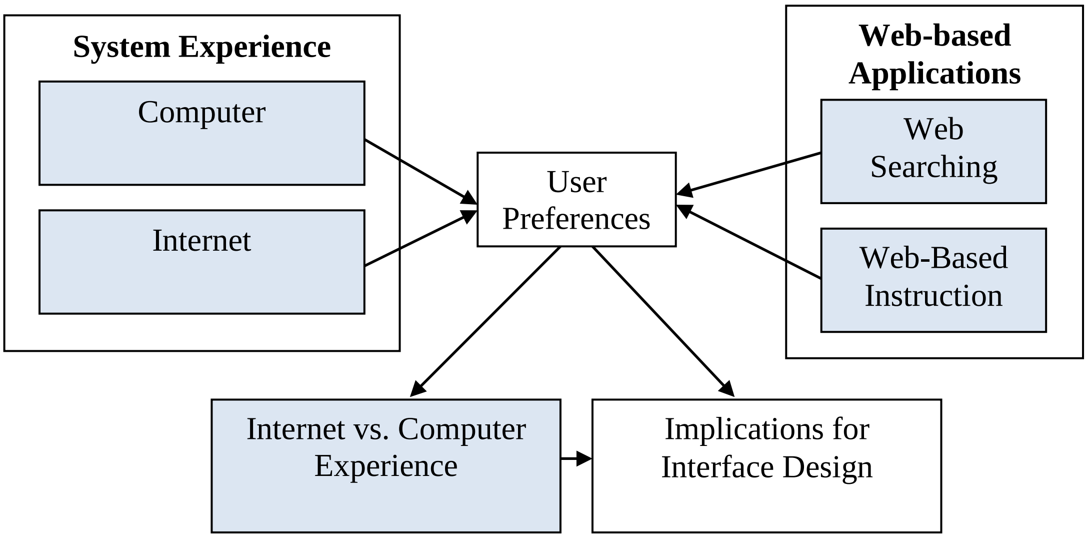



**Figure 4.1: Framework for Study 2**

## 4.2. Data Collection

The data collection and analysis procedures for the two datasets used in this thesis were described in the previous chapter (sections 3.2, 3.3 and 3.4). In this specific study, we introduced another human factor, system experience. This study focussed on two dimensions of system experience, including Internet experience and computer experience. Although the analysis procedure has already been described in the previous chapter, an additional step, a personal information sheet, was included within the data collection stage in order for the specific information needed for this study to be collected.

### 4.2.1 Personal Information Sheet

A personal information sheet was used to gather basic information of each participant, including gender, age, computer experience, and Internet experience. The data needed for Study 2 was collected from the last two of these sections. In order to identify participants' levels of both Internet and computer experience, they were presented with a 5-point Likert scale on which to rate their experience, with 1 being very little experience to 5, very experienced (Figure 4.5).

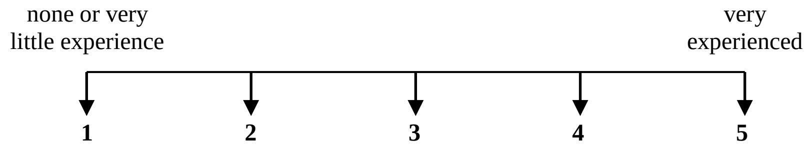



**Figure 4.2: Likert scale for rating Internet and computer experience**

#### 4.2.1.1 Dataset 1: Web Search Tools

As can be seen in Table 4.1, it seems that in general, students have more Internet experience than computer experience. These distributions vary from studies such as Schumacher and Morahan-Martin (2001), where students had more exposure to computers than with the Internet. In a time where the Internet is now a key part of education and research, perhaps users are more likely to have higher levels of Internet experience than computer experience. In addition, when looking at the participants' experience distributions (Table 4.1), it is clear that there is some form of link between the two types of experience. 35 out of the 37 participants who have a level 2 in computer experience have a level 2 in Internet experience. Likewise, 22 out of the 25 participants who have a level 3 in computer experience have a level 3 in computer experience, and 20 out of the 42 participants who have a level 4 in computer experience have a level 4 in Internet experience.

**Table 4.1. Dataset 1: Participants' Level of Experience**

| | | Computer Experience | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | |
| **Internet Experience** | **1** | 0 | 6 | 0 | 0 | 0 | **6** |
| | **2** | 2 | 35 | 0 | 0 | 0 | **37** |
| | **3** | 0 | 13 | 22 | 0 | 0 | **35** |
| | **4** | 0 | 5 | 17 | 20 | 0 | **42** |
| | **5** | 0 | 0 | 0 | 0 | 0 | **0** |
| **Total** | | **2** | **59** | **39** | **20** | **0** | **120** |

#### 4.2.1.2 Dataset 2: Web-Based Instruction Tools

Participants' experience distributions for the second dataset are very similar to those of the first in that they, in general, show that participants have higher levels of Internet experience than they do computer experience (Table 4.2). In particular, 19 of the 20 students who have a level 2 in Internet experience have a level 2 in computer experience. Similarly, 12 out of the 19 students who have a level 3 in Internet experience have a level 3 in computer experience, and 11 out of the 23 students who have a level 4 in Internet experience have a level 4 in computer experience. Having shown that both datasets had these same similarities in common, such a result suggests that the two types of experience could be associated.

**Table 4.2. Dataset 2 - Participants' Level of Experience**

| | | Computer Experience | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | |
| **Internet Experience** | **1** | 0 | 3 | 0 | 0 | 0 | **3** |
| | **2** | 1 | 19 | 0 | 0 | 0 | **20** |
| | **3** | 0 | 7 | 12 | 0 | 0 | **19** |
| | **4** | 0 | 3 | 9 | 11 | 0 | **23** |
| | **5** | 0 | 0 | 0 | 0 | 0 | **0** |
| | **Total** | **1** | **32** | **21** | **11** | **0** | **65** |

Even before looking at the analysis results, it is clear just from these two tables (Tables 4.1 and 4.2) that there is some similarity in the distributions of the two participant populations. For example, we can see that in general the students have more Computer Experience and roughly even Internet Experience.

## 4.3 Data Analysis

The data analysis procedure employed for Study 1 (described in the previous chapter: Section 3.4) was used again for this study. As shown in Figure 4.3, the data analysis procedure included two phases in order to identify the relationship between Internet experience and computer experience.

## 4.4 Findings – Dataset 1: Web Search Tools

As with the first study in this thesis, before beginning the data mining analysis a simple Pearson's correlation was used to see whether there was an initial relationship between the two types of experience. For this dataset, the test revealed that there is indeed a statistically significant (p=0.00) relationship between Internet and Computer

Experience. With this in mind, the data mining analysis procedure was completed and the results for the Web Searching dataset are discussed in the following sections.

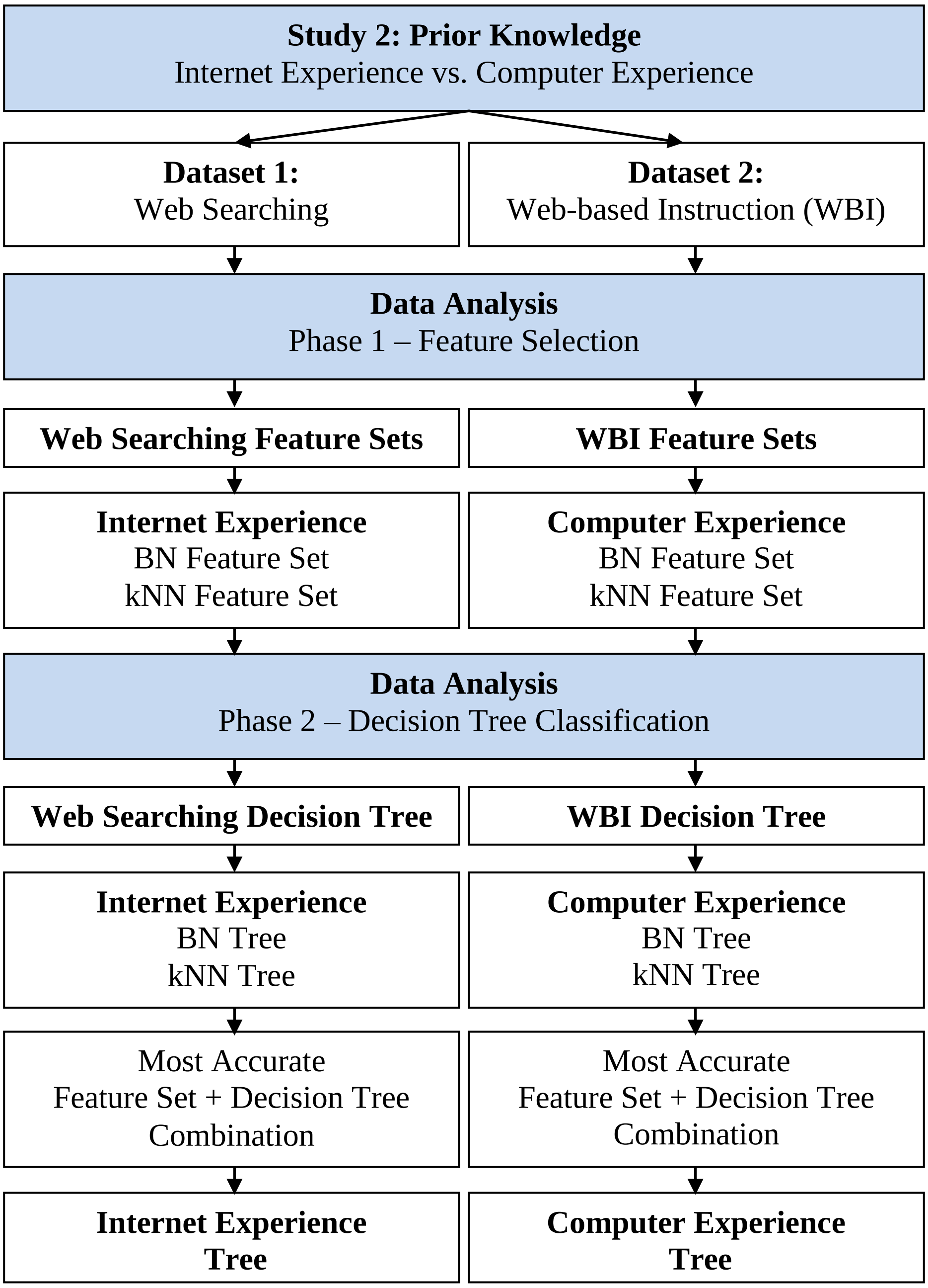



**Figure 4.3. Data Analysis Framework for Study 2**

### 4.4.1 Feature Selection

#### 4.4.1.1 Internet Experience

The BN and KNN classifiers selected 10 and 17 relevant features respectively for Internet Experience. When the two feature sets are compared, there are two features that were commonly selected by both.

- Q15 (*“Relevant subject content appears on the same page.”*)
- Q30 (*“Online instructions appear in a consistent location across screens.”*)

**Table 4.3. Dataset 1: Number of Features Selected per Family**

<table>
<tr><th rowspan="2">Experience</th><th colspan="4">Feature Sets</th></tr>
<tr><th>Classifier Family</th><th>Classifier</th><th># of Features Selected</th><th># of Features in Final Set</th></tr>
<tr><td rowspan="6">Internet Experience</td><td rowspan="3">BN</td><td>BNC</td><td>7</td><td rowspan="3">10</td></tr>
<tr><td>NB</td><td>7</td></tr>
<tr><td>AODE</td><td>3</td></tr>
<tr><td rowspan="3">KNN</td><td>NN</td><td>9</td><td rowspan="3">17</td></tr>
<tr><td>KNN</td><td>9</td></tr>
<tr><td>K*</td><td>3</td></tr>
<tr><td rowspan="6">Computer Experience</td><td rowspan="3">BN</td><td>BNC</td><td>42</td><td rowspan="3">27</td></tr>
<tr><td>NB</td><td>38</td></tr>
<tr><td>AODE</td><td>34</td></tr>
<tr><td rowspan="3">KNN</td><td>NN</td><td>25</td><td rowspan="3">24</td></tr>
<tr><td>KNN</td><td>20</td></tr>
<tr><td>K*</td><td>27</td></tr>
</table>

#### 4.4.1.2 Computer Experience

The BN and KNN classifiers selected 27 and 24 relevant features respectively to computer experience. When compared, the two feature sets have 16 features in common, which are listed below.

- Q2 (*“You can easily understand what options are available to you.”*)
- Q3 (*“There is some form of instruction to teach you how to do activity.”*)
- Q14 (*“The program avoids using computer jargon/technical terms.”*)
- Q15 (*“Relevant subject content appears on the same page.”*)
- Q28 (*“Menus with multiple levels have a mechanism that allows you go back.”*)
- Q29 (*“Error messages appear in the same location throughout the system.”*)

- Q30 (*"Online instructions appear in a consistent location across screens."*)
- Q31 (*"Icons are clearly labelled."*)
- Q33 (*"Heavy use of all uppercase letters on a webpage has been avoided."*)
- Q38 (*"Vertical and horizontal scrolling is possible in each window."*)
- Q42 (*"Messages are constructive, without overt or implied criticism to you."*)
- Q48 (*"Error messages let you know the cause of the problem."*)
- Q56 (*"The buttons that can cause serious consequences are located far away from low-consequence and high-use keys."*)
- Q67 (*"Different colours are used to discriminate links that you have visited from links that you have not visited."*)
- Q76 (*"Icons of different groups are visually distinctive."*)
- Q77 (*"Only information essential to the task is displayed on the screen."*)

#### 4.4.1.3 Internet Experience vs. Computer Experience

If the two sets of commonly selected features for Internet and Computer experience are compared, it is clear to see that two features are common in both sets. Q15 ("*Relevant subject content appears on the same page*.") and Q30 ("*Online instructions appear in a consistent location across screens.*"). These two features are concerned with content presentation and online help, which suggests that these are key factors in the relationship between the two dimensions of Internet and computer experience.

### 4.4.2 Decision Tree Classification

#### 4.4.2.1 Internet Experience

The algorithm that produced the highest classification accuracy was the C4.5 algorithm, with the most accurate feature set being the BN feature set with 100% (Table 4.4). The tree created with this combination can be found in Figure 4.4.

**Table 4.4. Internet Experience Classification Accuracy (%)**

| **Feature Set** | **Decision Tree Algorithm** | | |
|---|---|---|---|
| | C4.5 | CN2 | CART |
| BN | **100** | 90 | 95 |
| KNN | **98.667** | 90 | 95 |
| **Total Average** | **99.564** | 90 | 95 |

It is clear on inspection of Figure 4.4 that users with lower levels of Internet experience do not think it important to have short, descriptive titles. Conversely, users with higher levels of Internet experience think it important to have relevant content appearing on the same page and they think it important to have clear, descriptive titles. This could be because the former lack experience of using Internet applications. Thus, they do not know that there are a variety of approaches, i.e. short and descriptive page or section titles, which allow them to get a preliminary view of the subject content. On the other hand, the latter have more experience so they expect to locate information efficiently using one or more of these approaches.

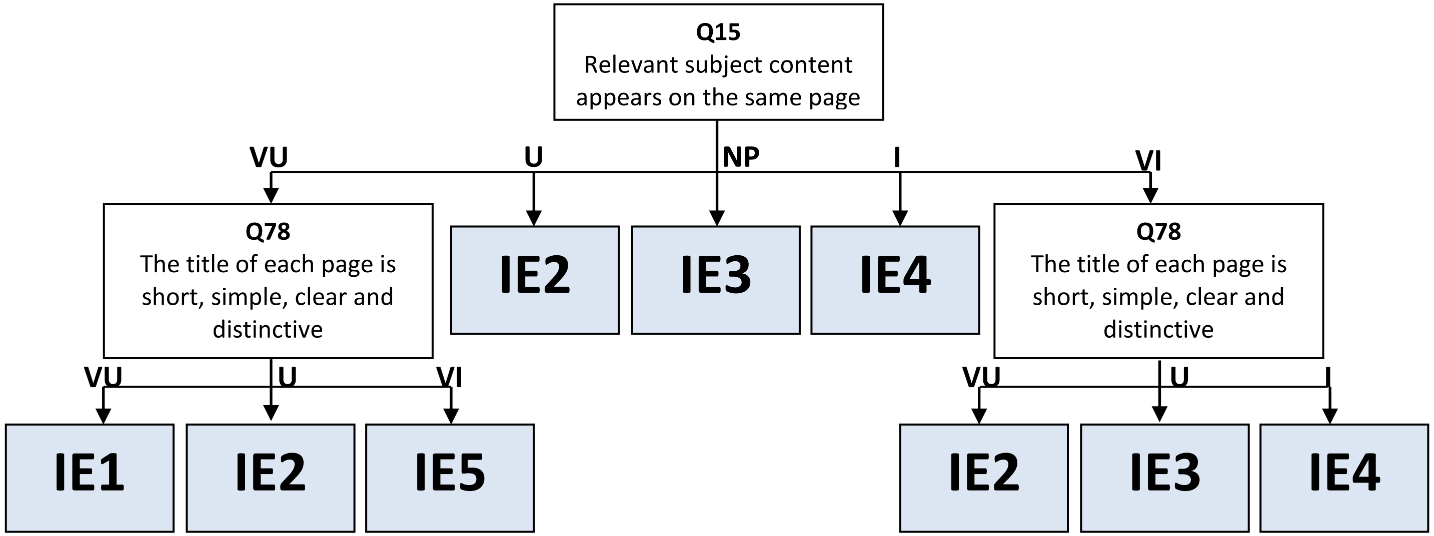



**Figure 4.4. Dataset 1 - Internet Experience Decision Tree**

**Table 4.5. Decision Rules for the Internet Experience Decision Tree**

| Dataset | IE Level | Decision Rules |
|---|---|---|
| Dataset 1: Web Searching | IE 1 | If users think it very unimportant for relevant subject content to appear on the same page, AND they think it very unimportant that menu items are brief yet long enough to describe the subject content, they have an IE level of 1. |
| | IE 2 | If users think it unimportant for relevant subject content to appear on the same page, they have an IE level of 2. |
| | IE 2 | If users think it very unimportant for relevant subject content to appear on the same page, AND they think it unimportant that menu items are brief yet long enough to describe the subject content, they have an IE level of 2. |
| | IE2 | If users think it very important for relevant content to appear on the same page and they think it unimportant that the title of each page is short, simple, clear and distinctive, they have an IE level of 2. |
| | IE 2 | If users think it very important for relevant content to appear on the same page and they think it important that the title of each page is short, simple, clear and distinctive, they have an IE level of 2. |
| | IE 3 | If users have no preference for relevant subject content to appear on the same page, they have an IE level of 3. |
| | IE3 | If users think it very important for relevant content to appear on the same page and they think it unimportant that the title of each page is short, simple, clear and distinctive, they have an IE level of 3. |
| | IE 4 | If users think it important for relevant subject content to appear on the same page, they have an IE level of 4. |
| | IE 4 | If users think it very unimportant for relevant subject content to appear on the same page, AND they think it important that menu items are brief yet long enough to describe the subject content, they have an IE level of 4. |
| | IE 5 | If users think it very unimportant for relevant subject content to appear on the same page, AND they think it very important that menu items are brief yet long enough to describe the subject content, they have an IE level of 5. |

#### 4.4.2.2 Computer Experience

The algorithm that classified the feature sets with the highest overall accuracy was the CART algorithm, with the BN feature set classifying with the highest accuracy (99.5763%). The resulting tree from this combination can be found in Figure 4.5.

**Table 4.6. Computer Experience Classification Accuracy (%)**

| Feature Set | **Decision Tree Algorithm** | | |
|---|---|---|---|
| | *C4.5* | *CN2* | ***CART*** |
| BN | 94.6667 | 85 | **99.5763** |
| KNN | 96.6667 | 85 | **98.5412** |
| Total Average | 95.333 | 85 | **99.0154** |

It is clear from this tree (Figure 4.5) that users with lower levels of computer experience do not think it is important to have online instructions or error messages appearing in the same place, whereas users with higher levels prefer them to be in the same place throughout the system. Many computer applications display the online instructions or error messages consistently in the same place, therefore the latter will have experience of this and expect it in a computer application and they know that this can help them to find online instructions and error messages easily. On the other hand, the former will not have experience of using other computer applications which put online instruction and error messages in this way. Therefore, they will not expect online instructions or error messages to be consistently located in the same place so they will think this less important.

Users with lower levels of computer experience do not think it important to have clearly labelled icons or relevant information appearing on the same page. Conversely, users with higher levels of computer experience think it important to have clearly labelled icons, relevant information appearing on the same page and to avoid heavy use of uppercase characters. This is probably because users with higher levels of computer experience will have known that such things can help them to complete their tasks more efficiently, whereas users with lower levels of computer experience will not know such value.

Users with higher levels of computer experience think that it is important to be able to customise the interface (e.g. change the font size or colour). This may be because they are more familiar within their environment and have already discovered their personal preferences that will help them to complete their tasks more efficiently. On the other hand, users with lower levels of computer experience will not have as much experience and will be more concerned with learning how to complete their task correctly, rather than completing it in a more efficient way. In this way, users with lower levels of computer experience will need more instruction because unlike those with higher levels, they do not have previous knowledge that they can apply to the task.

#### 4.4.2.3 Internet Experience vs. Computer Experience

Overall, it is clear that users with higher levels of Internet or computer experience are more proficient at identifying which features can help them to complete their tasks more quickly efficiently. For example, users with high levels of Internet experience think it important to have relevant content appearing on the same page and pages labelled with clear, descriptive titles. For computer experience, users with high levels also think it important to have relevant content appearing on the same page. In addition, they think it important for error messages and online instructions to be displayed in the same location throughout the system, clearly labelled icons and the ability to customise the interface environment.

In contrast, those users with lower levels of Internet or computer experience have difficulty identifying which features can help them to complete their tasks more quickly and efficiently, instead focusing on learning how to complete their task correctly. For example, users with low levels of computer experience think it important to easily and clearly able to tell which icon is selected. This feature helps them to know which icon is selected, which can help them complete the task in the correct way, but does not necessarily help them complete their task more efficiently. Thus, users with lower levels of experience will need more instruction and support in firstly, learning how to complete their task correctly, and secondly, identifying the features which will help them to complete their task more efficiently.

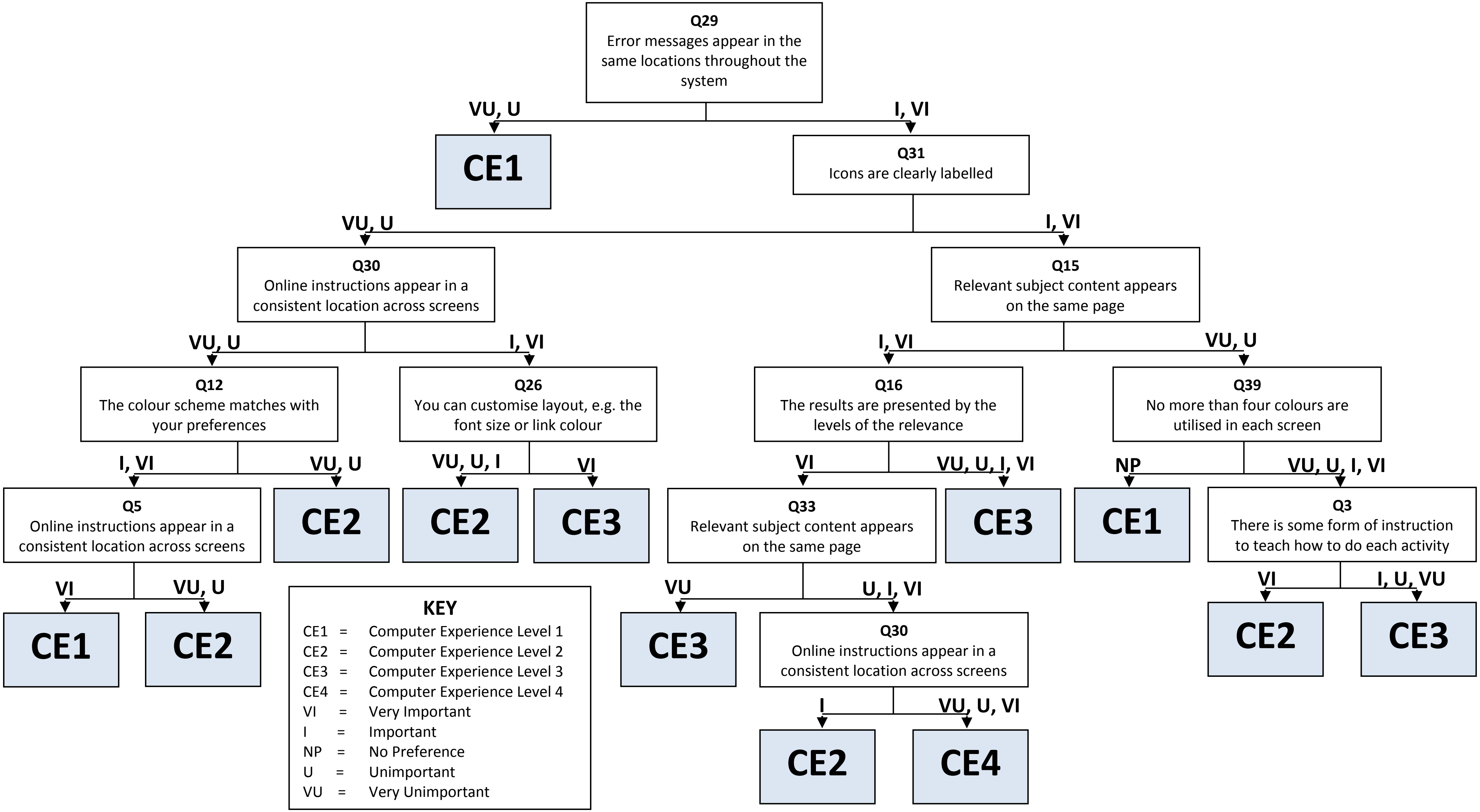



**Figure 4.5. Dataset 1 – Computer Experience Decision Tree**

**Table 4.7. Decision Rules for Computer Experience Tree**

| Dataset | Experience Level | Decision Rules |
|---|---|---|
| Dataset 1: Web Searching | CE 1 | If users think it is very unimportant or unimportant for error messages to appear in the same location throughout the system, they have a CE level of 1. |
| | CE 1 | If users think it is very important or very important for error messages to appear in the same location throughout the system, AND very unimportant or unimportant for online instructions to appear in the same location throughout the system, AND it very important or very important that the colour scheme matches with their preferences, AND think it important that it is easy to understand which icon has been selected, they have a CE level of 1. |
| | CE 1 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND very unimportant or unimportant for relevant content to appear on the same page, AND have no preference about whether no more than four colours are used on the screen, they have a CE level of 1. |
| | CE 2 | If users think it is very important or very important for error messages to appear in the same location throughout the system, AND very unimportant or unimportant for online instructions to appear in the same location throughout the system, AND it very unimportant or unimportant that the colour scheme matches with their preferences, they have a CE level of 2. |
| | CE 2 | If users think it is very important or very important for error messages to appear in the same location throughout the system, AND very unimportant or unimportant for online instructions to appear in the same location throughout the system, AND it very important or very important that the colour scheme matches with their preferences, AND think it is not important that it is easy to understand which icon has been selected, they have a CE level of 2. |
| | CE 2 | If users think it is very important or very important for error messages to appear in the same location throughout the system, AND very unimportant or unimportant for online instructions to appear in the same location throughout the system, AND do not think it very important to be able to customise the layout , e.g. font size and colour, they have a CE level of 2. |
| | CE 2 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND important or very important for relevant content to appear on the same page, AND very important that the results are presented by level of relevance, AND not very unimportant that heavy use of all uppercase letters is avoided, AND important that online instructions appear in the same location throughout the system, they have a CE level of 2. |
| | CE 2 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND very unimportant or unimportant for relevant content to appear on the |

| Dataset | Experience Level | Decision Rules |
|---|---|---|
| | | same page, AND have any preference about whether no more than four colours are used on the screen, AND it very important that there is some form of instruction to teach you how to do each activity, they have a CE level of 2. |
| | CE 3 | If users think it is very important or very important for error messages to appear in the same location throughout the system, AND very unimportant or unimportant for online instructions to appear in the same location throughout the system, AND think it very important to be able to customise the layout , e.g. font size and colour, they have a CE level of 3. |
| | CE 3 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND important or very important for relevant content to appear on the same page, AND not very important that the results are presented by level of relevance, they have a CE level of 3. |
| | CE 3 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND important or very important for relevant content to appear on the same page, AND very important that the results are presented by level of relevance, AND very unimportant that heavy use of all uppercase letters is avoided, they have a CE level of 3. |
| | CE 3 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND very unimportant or unimportant for relevant content to appear on the same page, AND have any preference about whether no more than four colours are used on the screen, AND it not very important that there is some form of instruction to teach you how to do each activity, they have a CE level of 3. |
| | CE 4 | If users think it is very important or important for error messages to appear in the same location throughout the system, very important or important for icons to be clearly labelled, AND important or very important for relevant content to appear on the same page, AND very important that the results are presented by level of relevance, AND not very unimportant that heavy use of all uppercase letters is avoided, AND not important that online instructions appear in the same location throughout the system, they have a CE level of 4. |

### 4.4.3 Statistical Analysis

An ANOVA was completed to discover the features that were significant to both Internet and computer experience. For the Web Searching dataset, the two features (Q15 and Q78) that were highlighted on the Internet experience decision tree were both found to be statistically significant (p=<0.05), whereas only 4 out of the 10 features (Q15, Q30, Q31 and Q33) highlighted on the computer experience decision tree were found to be statistically significant.

### 4.5 Findings – Dataset 2: Web-Based Instruction Tools

In the first stage of analysis, six classifiers from two different families were used to select two relevant feature sets for each Internet Experience and computer experience. Subsequently, the sets for Internet experience were compared with those for computer experience before being taken on to the decision tree classification stage.

### 4.5.1 Feature Selection

#### 4.5.1.1 Internet Experience

The Bayesian Network (BN) family of classifiers selected five features as relevant to Internet experience, whereas the k-Nearest Neighbour (KNN) family of classifiers selected only four (Table 4.8). Of these two feature sets, three features were commonly selected: Q12 (‘*The map in this tutorial gives a meaningful framework of HTML*’), Q14 (‘*After using this system I can I can easily use my knowledge to design homepages*’) and Q15 (‘*I found it hard to select relevant information using the map*’). This suggests that the map is related to Internet experience. This is probably because the map is beneficial to novices, who have lower levels of Internet experience and easily get lost. The map can help them incorporate the document structure into the conceptual structure (Dee-Lucas & Larkin, 1995) so that they can effectively integrate individual topics (Potelle & Rouet, 2003). Thus, there is a need to provide effective maps to accommodate students with different levels of Internet experience.

#### 4.5.1.2 Computer Experience

The BN family of classifiers selected five features as relevant to computer experience, whereas the KNN family of classifiers selected only three (Table 4.8). Of these two feature sets, three features were commonly selected: Q1 (‘*It is difficult to learn the basics of HTML using this tutorial without the help of a person*’), Q14 (‘*After using this system I can easily use my knowledge to design homepages*’) and

Q15 (‘*I found it hard to select relevant information using the map*’). Q1 and Q14 relate to how confident students are in applying the knowledge they have learnt. This would suggest that students’ confidence in using the system is greatly linked to the level of computer experience, which refers to the amount and type of computer skills a person acquires over time (Howard and Smith, 1986). This finding echoes the claims made by Smith *et al.* (1999), which indicate that computer experience is positively associated with the students’ attitudes.

**Table 4.8: Dataset 2: Number of Features Selected per Family**

| **Experience** | **Feature Sets** | | | |
|---|---|---|---|---|
| | **Classifier Family** | **Classifier** | **# of Features Selected** | **# of Features in Final Set** |
| **Internet Experience** | BN | BNC | 4 | 5 |
| | | NB | 4 | |
| | | AODE | 5 | |
| | KNN | NN | 10 | 4 |
| | | KNN | 5 | |
| | | K* | 5 | |
| **Computer Experience** | BN | BNC | 7 | 5 |
| | | NB | 6 | |
| | | AODE | 7 | |
| | KNN | NN | 3 | 3 |
| | | KNN | 7 | |
| | | K* | 5 | |

#### 4.5.1.3 Internet Experience vs. Computer Experience

When comparing the findings for both Internet and computer experience, it is clear to see that there are two features in common: Q14 (‘*After using this system I can easily use my knowledge to design homepages*’) and Q15 (‘*I found it hard to select relevant information using the map*’). These two features relate to the way that students find information in an environment and then apply the knowledge learnt. Computer experience covers a broader range of skills while Internet experience focuses on specific skills. Regardless of the broad range of computing skills or specific Internet skills, these skills are related to the ways that students are able to find and apply knowledge in a system. Therefore, there is a need to consider these two pertinent skills together when we design facilities for students to find information in the WBI tool.

### 4.5.2 Decision Tree Classification

Once the feature sets were identified, decision tree classification was then performed to find the most accurate feature set for both Internet and computer experience, which were then used to build decision trees and used to illustrate users' preferences of WBI tools.

#### 4.5.2.1 Internet Experience

As stated in the previous section, three algorithms were used to initially test the accuracy of the feature sets. The classification accuracies were then recorded and the most accurate algorithm/feature set combination was used to build the decision tree that would illustrate users' preferences. For Internet Experience, the CART algorithm and the KNN feature set provided the highest accuracy (88.2%) (Table 4.9). Therefore, this combination was used to build the final decision tree (Figure 4.6).

**Table 4.9. Internet Experience Classification Accuracy (%)**

| Feature Set | **Decision Tree Algorithm** | | |
|---|---|---|---|
| | *C4.5* | *CN2* | ***CART*** |
| BN | 66.15 | 58.4615 | **81.3218** |
| KNN | 61.5385 | 56.9231 | **88.24405** |
| Total Average | 63.84425 | 57.6923 | **84.78293** |

It is clear from the Internet Experience decision tree (Figure 4.6) that those students with lower levels of experience found it difficult to apply the knowledge they learnt in the tutorial to designing homepages. Conversely, those students with higher levels of experience found it easier to apply the knowledge and design homepages by themselves. This suggests that the higher a student's level of Internet experience, the easier they will find it to apply their knowledge to the system, which in this case, was a tutorial that aimed to teach the student how to use HTML. Furthermore, this suggests that Internet experience is related to designing Web pages, as the more

Internet experience a student has, the easier they will be able to design homepages. This would also explain why the students with lower levels of experience agree that more examples are needed to help them understand the content, differing from those with higher levels of experience who disagree. Because of their different previous experience, the latter were more confident in applying the knowledge from the tutorial, whereas the former felt they needed more examples to learn to apply the knowledge. This implies that examples, which are down-to-earth visual materials (Chen and Liu, 2008), are useful facilities to support students with lower levels of Internet experience. Additionally, those students with higher levels of Internet experience found it easier than those with lower levels of Internet experience to select relevant information using the map. A map is the standard way of presenting information on a Web page so users with high levels of experience will have encountered this before and be familiar with finding information in this way. In comparison, those with less experience in using the Internet would not be so familiar with the map so they may struggle to find information using this feature.

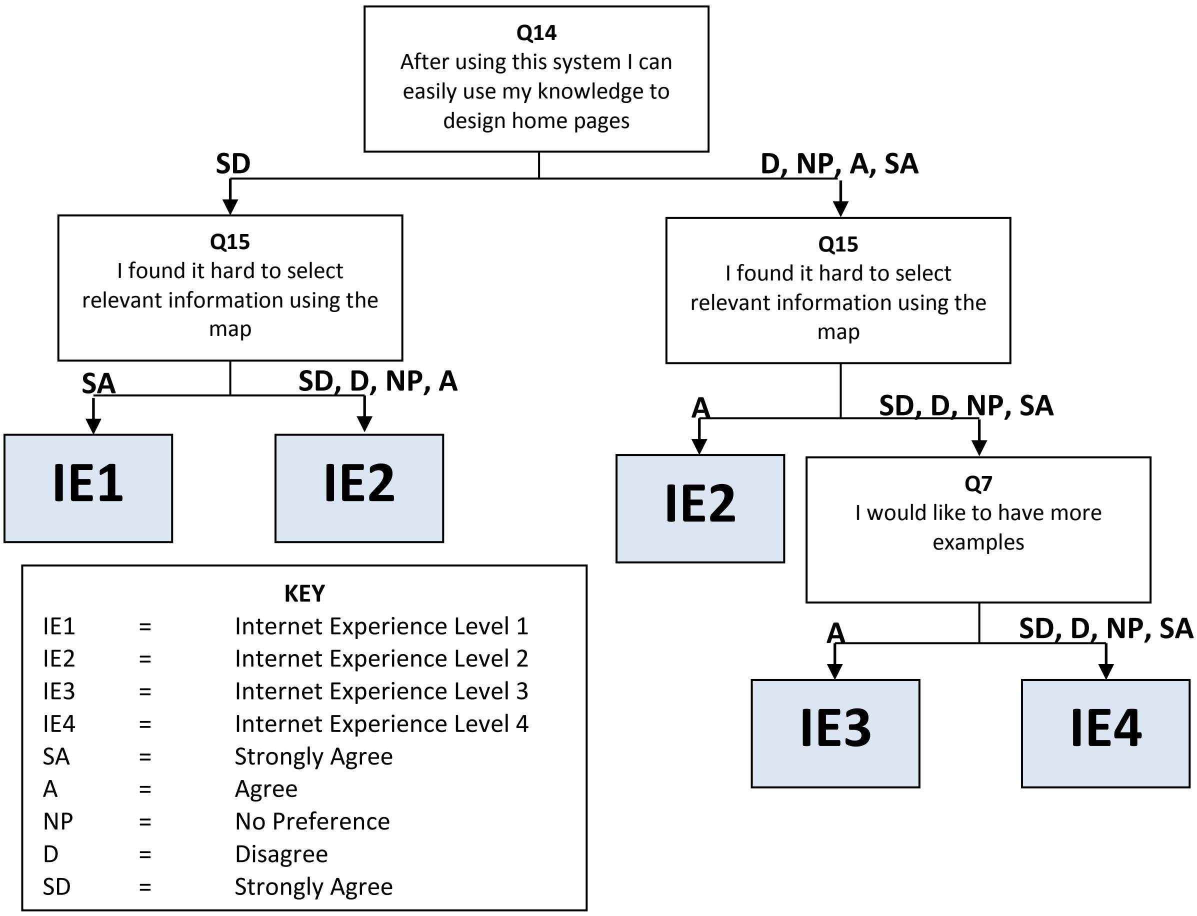



**Figure 4.6: Dataset 2 - Internet Experience Decision Tree**

**Table 4.10. Decision Tree Rules for Internet Experience Tree**

| Dataset | IE Level | Decision Rules |
|---|---|---|
| Dataset 2: Web-based Instruction | IE 1 | If users strongly disagree that after using this system they can easily use their knowledge to design homepages, AND they strongly agree that they it is hard to select relevant information using the map, they have an IE level of 1. |
| | IE 2 | If users strongly disagree that after using this system they can easily use their knowledge to design homepages, AND they do not strongly agree that they it is hard to select relevant information using the map, they have an IE level of 2. |
| | IE 2 | If users do not strongly disagree that after using this system they can easily use their knowledge to design homepages, AND they agree that it is hard to select relevant information from the map, they have an IE level of 2. |
| | IE 3 | If users do not strongly disagree that after using this system they can easily use their knowledge to design homepages, AND they do not agree that it is hard to select relevant information from the map, AND they agree that they would like to have more examples, they have an IE level of 3. |
| | IE 4 | If users do not strongly disagree that after using this system they can easily use their knowledge to design homepages, AND they do not agree that it is hard to select relevant information from the map, AND they do not agree that they would like to have more examples, they have an IE level of 4. |

#### 4.5.2.2 Computer Experience

For computer experience, both the BN and KNN feature sets scored the highest accuracy (94.4%) when classified using the C4.5 algorithm (Table 4.11). Therefore, this combination was assumed to most accurately illustrate students' preferences based on their computer experience (Figure 4.7).

**Table 4.11. Computer Experience Classification Accuracy**

| Feature Set | **Decision Tree Algorithm** | | |
|---|---|---|---|
| | ***C4.5*** | *CN2* | *CART* |
| BN | **94.40218** | 87.6923 | 89.2308 |
| KNN | **94.40218** | 86.1538 | 89.2305 |
| Total Average | **94.40218** | 86.92305 | 89.23065 |

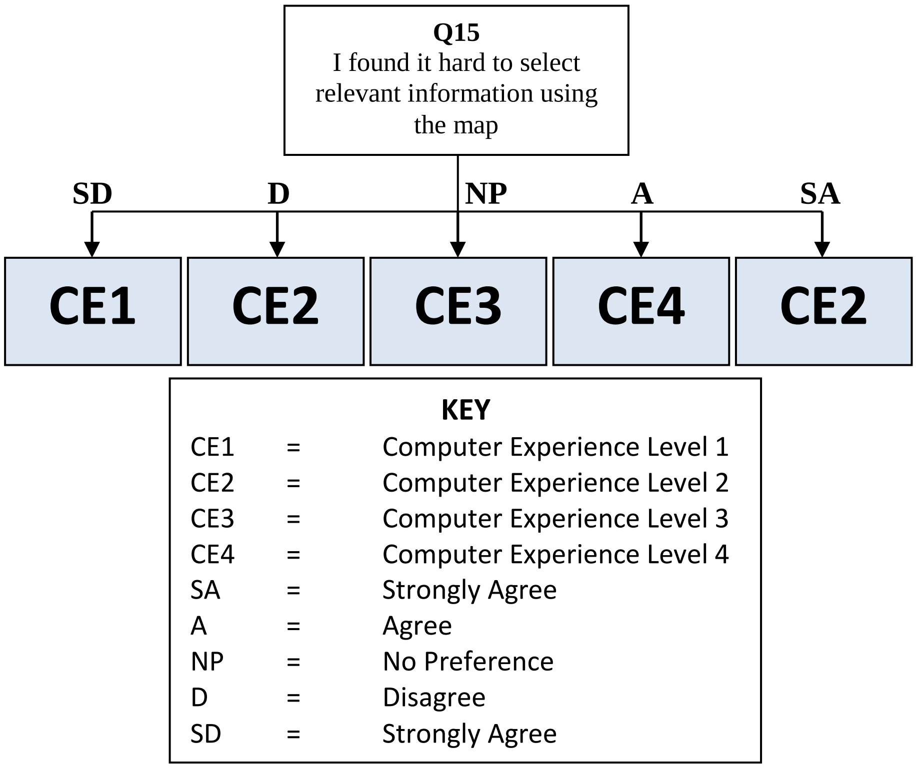



**Figure 4.7: Dataset 2 - Computer Experience Decision Tree**

**Table 4.12. Decision Tree Rules for Computer Experience Decision Tree**

| Dataset | CE Level | Decision Rules |
|---|---|---|
| Dataset 2: Web-based Instruction | CE1 | If users strongly disagree that it is hard to select relevant information using the map, they have a CE level of 1. |
| | CE2 | If users disagree that it is hard to select relevant information using the map, they have a CE level of 2. |
| | CE3 | If users have no preference about whether it is hard to select relevant information using the map, they have a CE level of 3. |
| | CE4 | If users agree that it is hard to select relevant information using the map, they have a CE level of 4. |
| | CE2 | If users strongly agree that it is hard to select relevant information using the map, they have a CE level of 2. |

As this tree shows (Figure 4.7), it was possible to classify students into their corresponding computer experience levels through the use of just one feature, Q15. The classification accuracy is very high (94.402%), so this signifies that Q15 is an extremely important consideration when considering computer experience.

The tree shows that, generally, the more computer experience students have, the harder they find it to select relevant information using the map. Conversely, the less experience they have, the easier it is for them to select relevant information on the map. Although the majority of Internet applications present information using a map, other computer applications, such as emails or word processing, do not necessarily present information in this way. Even though a student might have a lot of experience in using computer applications, they will more than likely have difficulty in applying this previous experience to using Internet applications because the environments are so different. For example, whilst a user that frequently uses a keyboard to type emails or use the DVD drive to burn movies, this knowledge will not help them when it comes to selecting relevant information from a map in the WBI tools. This would make it difficult for students with high levels of computer experience to apply their prior knowledge into an Internet environment because they have learnt to find relevant information in a different manner. On the other hand, students with less computer experience will approach the task of finding relevant information on the map afresh, without trying to apply any prior knowledge that might make them feel less frustrated with using the map.

There is one contradiction to the above finding that students with the second level of computer experience had diverse preferences. Some of them strongly agreed it is difficult to find information using a map whilst others disagreed with this statement. A possible reason for those who showed disagreement was that their previous experience was more than complete novices, but less than those who use computer applications on a more regular basis. Another words, they had a small amount of knowledge about computer applications. Because the level of this previous knowledge and experience was quite low, they might not be able to apply this knowledge successfully to the new Internet environment. Perhaps this was because they had just enough previous experience with using computer applications to try to transfer this knowledge to the new environment, but became confused when it was clear that the applications were too different and did not have the advanced skills necessary to work it out for themselves.

#### 4.5.2.3 Internet Experience vs. Computer Experience

It is clear that when the Internet Experience decision tree (Figure 4.6) and the computer experience decision tree (Figure 4.7) are compared, there is one particular feature that both have in common: Q15 (‘*I found it hard to select relevant information using the map*’). The trees show that the more Internet experience students have, the easier it was for them to find relevant information using the map structure. However, this does not necessarily mean that the more computer experience students have, the easier it is for them to find information using the map structure. For example, students with high levels of computer experience found it difficult to use the map to find relevant information in the tutorial. This suggests that computer experience may not be comparable to Internet experience and students would have difficulty in applying their Computer knowledge in an Internet environment, such as the WBI tool used in this study. On the other hand, this finding also reveals that the map may not be suitable to all of students, for example students with lower levels of Internet experience or those with higher levels of computer experience. Therefore, only providing maps in WBI tools cannot accommodate students’ diverse preferences. More specifically, the WBI should provide multiple navigation facilities to cater for their individual differences, e.g., main menu or alphabetical index.

There is another question, Q14 (‘*After using this system I can easily use my knowledge to design homepages*’), that was highlighted in the Internet experience tree and in the Feature Selection results (Section 4.1.3). However, it did not appear in the computer experience tree. On closer inspection, Q14 is related to applying knowledge the student has learnt in the system. In this case, the system is a WBI tool. WBI tools are on the Internet, which may explain why Q14 is more related to Internet Experience than to computer experience. Thus, there is a need to consider the context of the applications when we want to provide personalisation for the applications based on a particular human factor. In this study, the context is a Web-based tool so Internet experience plays a more important role.

### 4.5.3 Statistical Significance

For this dataset, only 1 (Q15) out of the 3 (Q7, Q14, Q15) features highlighted on the Internet experience tree was found to be statistically significant. In addition, the one feature (Q15) highlighted on the computer experience tree was found to be statistically significant.

## 4.6 Findings: Dataset 1 vs. Dataset 2

### 4.6.1 Internet Experience

If the trees for Internet experience from the two datasets are compared, it is clear that the more Internet experience a user has, the easier it is for them to complete their tasks with more speed and efficiency. Firstly, not only can they fully utilise the menu structure to find relevant information, but it is easy for them to identify which features will help them to achieve this goal. For example, in dataset 1, the more Internet experience a user has, the easier it is to identify which features can help them complete their tasks more quickly and efficiently. In dataset 2, the more Internet Experience a user has the easier it is for them to find information using the map.

### 4.6.2 Computer Experience

If the trees for computer experience from the two datasets are compared, it is clear that lower levels of computer experience face difficulties in identifying which features can help them to complete their tasks more efficiently and quickly. In addition, it is also highlighted in the second dataset that there is a contradiction to the assumed rule that computer experts are like Internet experts. In Dataset 2, computer experience experts are shown to find it hard and be as confused as novices when it comes to finding information on the map. The map is a unique feature of Internet applications and so it is likely that the user will find this method of navigation hard even though they might have previous experience with other computer applications. Such a finding suggests that the skills learnt in a computer environment are different to those in an Internet environment, and thus not completely comparable.

### 4.6.3 Internet vs. Computer Experience

Comparing the two datasets, it is clear that, unlike Internet Experience, a higher level of computer experience does not necessarily mean that users will be able to complete their tasks within a web-based application environment quickly and efficiently. Perhaps this is because the term ‘Computer Experience’ (CE) covers a wide area, stretching from burning a DVD to coding a computer program, whereas the term ‘Internet Experience’ (IE) is related to a very narrow set of applications in comparison. As such, a user may have high levels of experience in using computer applications, but their skills may not be transferable to Internet applications, simply because the environments are so different. However, although those users with more computer experience will find it difficult to apply their previous knowledge to an Internet environment, the results of this study show that they are able to identify which features will help them to complete their tasks more efficiently. For instance, those users with more computer experience know that clear, descriptive titles will help them to locate relevant information in a much more efficient manner. This suggests that although IE novices, CE novices and CE experts need additional support when using an Internet application, the two former will need more support in identifying which features will help them complete their tasks with more efficiency. For example, by highlighting the fact that online help messages appear in the same location throughout the system, the user will be able to get to the root of the problem quicker for knowing what the problem is and be able to find a solution more quickly. On the other hand, CE experts may encounter less difficulty in identifying such features than IE novices, CE novices.

### 4.7 Summary

This study investigated how a users' level of both Internet and computer experience influenced their preferences for Web-based applications. In particular, the relationship between Internet and computer experience was evaluated, concluding that although there are some similarities between the two dimensions of experience, the two are not equally alike. This is shown in particular by the results in Dataset 2, where computer experts and Internet novices displayed the same preferences. The results are summarised in Figure 4.8.

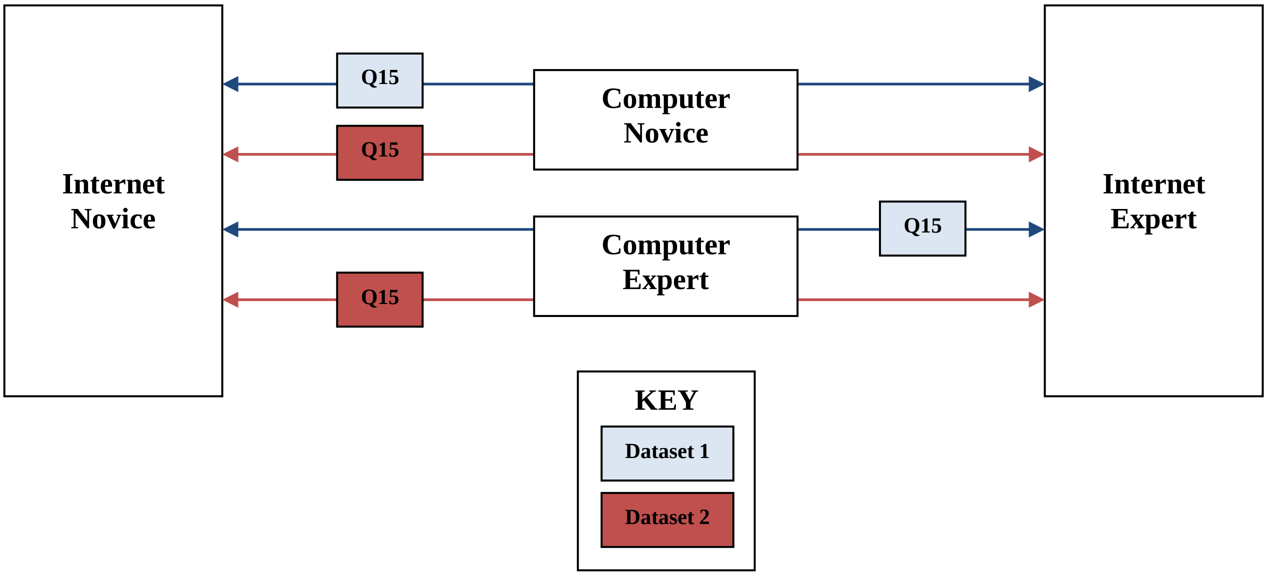



**Figure 4.8. Relationship between Internet Experience and Computer Experience**

# Chapter 5: Study 3 – Cognitive Style vs. Previous Experience vs. Gender Differences

## 5.1 Introduction

This chapter will describe the final of the three studies examined within this thesis. The third study combines the results from the first two studies, which examined cognitive style and system experience, with a new set of results, gender differences, with the aim of discovering their overall effects on user preferences of Web-based applications.

As users have many human factors, it can be assumed that more than one of these factors will affect users' preferences at any one time. Some works have suggested that there are some similarities in the relationship between the factors examined in this thesis. For example, Novices, Females and Field Dependent users are purported to display the same behaviour and preferences. Likewise, this is supposedly the same for Experts, Males, and Field Independent users (Chen, 2000; Fan, 2005). As with the investigations between cognitive style and system experience, we have shown that just because they exhibit similar behaviours, it is not always correct to say that they are totally equal. Therefore, it is necessary to conduct an examination into the relationships between these three human factors to see how similar these types of user actually are. More specifically, the following research question is proposed as the aim of this particular study:

*To what extent is the similarity between:*
*a) Novices, Females, Field Dependent users and Holists, and*
*b) Experts, Males, Field Independent users and Serialists?*

Such an investigation will have a number of contributions. Firstly, the study will help to gain a deeper understanding of the differences in preference between male and female users. Secondly, if a relationship is found, it will help to inform the design of Web-based applications to make them more efficient to use for the different types of user, helping to ensure future success. Thirdly, it is becoming increasingly obvious that designers have to take into account more than one human factor when designing effective Web-based applications. Such an investigation will test the integrated data mining technique this thesis proposes to see whether it can capture the interaction between these human factors.

Following the framework shown in Figure 5.1, the chapter will start by presenting the results related to the third human factor; gender differences. Subsequently, a detailed analysis that compares the chosen decision trees for cognitive style, system experience and gender differences to identify the similarities and differences in preference between the different types of user. Finally, reflections and the implications for the interface design of Web-based applications will be discussed.

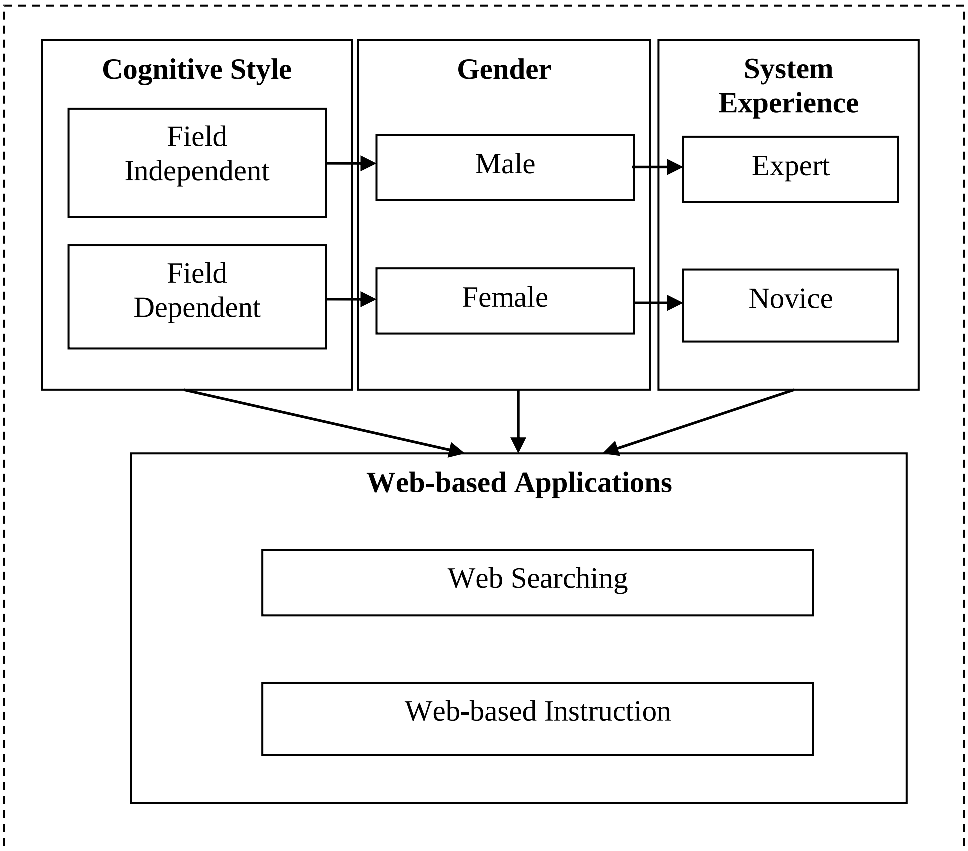



**Figure 5.1: Study 3 Framework**

## 5.2 Data Collection

### 5.2.1 Personal Information Sheet

The additional information needed for this study was obtained by a question that asked for the participant's gender in the personal information sheet. For Dataset 1, the participants were almost evenly divided among the sexes, with 61 males and 59 females. Likewise, participants' gender in Dataset 2 was evenly distributed with 32 males and 33 females (Figure 5.2).

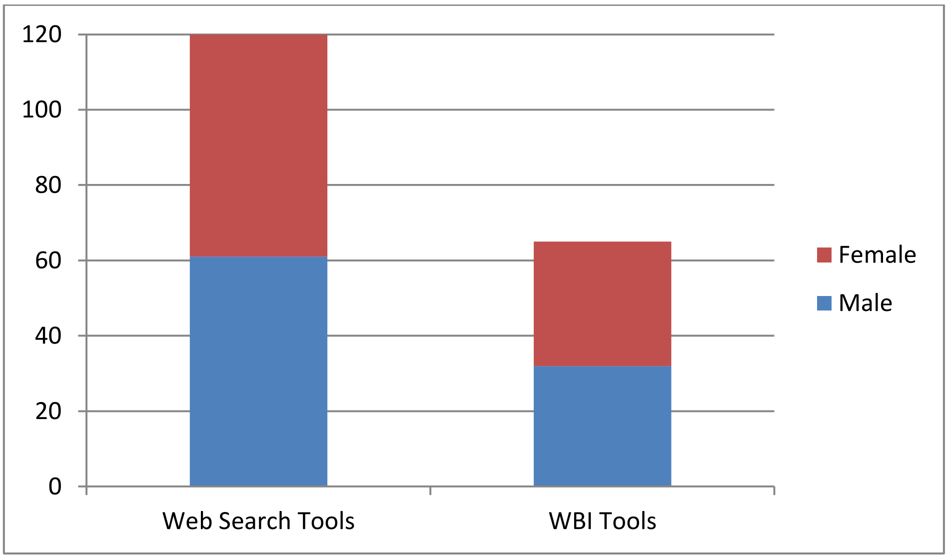



**Figure 5.2. Participants' Gender Distributions**

## 5.3 Results: Dataset 1 – Web Search Tools

As with the other two studies, before the data analysis for Study 3 commenced, a Pearson's correlation was run for both datasets to gain an initial idea of the relationship between gender and users' preferences. Unfortunately, neither of the datasets returned a significant result (where $p<0.05$). Subsequently, data analysis was completed, beginning with the feature selection.

### 5.3.1 Feature Selection

Of the original 90 features obtained from Study 2, the BN family of classifiers selected 19 as relevant to gender whereas the KNN family selected 26 (Table 5.1). Of these two feature sets, 12 features were commonly selected by both:

- Q2 (“*You can easily understand what options are available to you.*”)
- Q6 (“*There is always a way to get back to the home page from any pages.*”)
- Q9 (“*The names of links and buttons are understandable.*”)
- Q10 (“*High contrast colour scheme applied to present text/background.*”)
- Q11 (“*Menu items are arranged alphabetically.*”)
- Q12 (“*The colour scheme matches with your preferences.*”)
- Q13 (“*The length of the page matches with your expectation.*”)
- Q19 (“*From the menu list, you can see the relationships among the sub-topics presented in the menu.*”)
- Q21 (“*Each window has a title.*”)
- Q22 (“*Navigation buttons are labelled clearly and distinctively.*”)
- Q50 (“*Error messages avoid the use of exclamation points.*”)
- Q87 (“*You can easily switch between the help system and your work.*”)

### 5.3.2 Decision Tree Classification

When these two subsets were used to run the three decision tree algorithms, the BN feature set was found to be the most accurate using the C4.5 algorithm. Therefore, this classifier family/decision tree algorithm combination was used to build the decision tree that represents users’ preferences for Web search engines.

**Table 5.1: Dataset 1: Number of Features Selected per Family**

<table>
<tr><th rowspan="2"></th><th colspan="4">Feature Sets</th></tr>
<tr><th>Classifier Family</th><th>Classifier</th><th># of Features Selected</th><th># of Features in Final Set</th></tr>
<tr><td rowspan="6">Gender</td><td rowspan="3">BN</td><td>BNC</td><td>32</td><td rowspan="3">19</td></tr>
<tr><td>NB</td><td>24</td></tr>
<tr><td>AODE</td><td>29</td></tr>
<tr><td rowspan="3">KNN</td><td>NN</td><td>27</td><td rowspan="3">26</td></tr>
<tr><td>KNN</td><td>27</td></tr>
<tr><td>K*</td><td>19</td></tr>
</table>

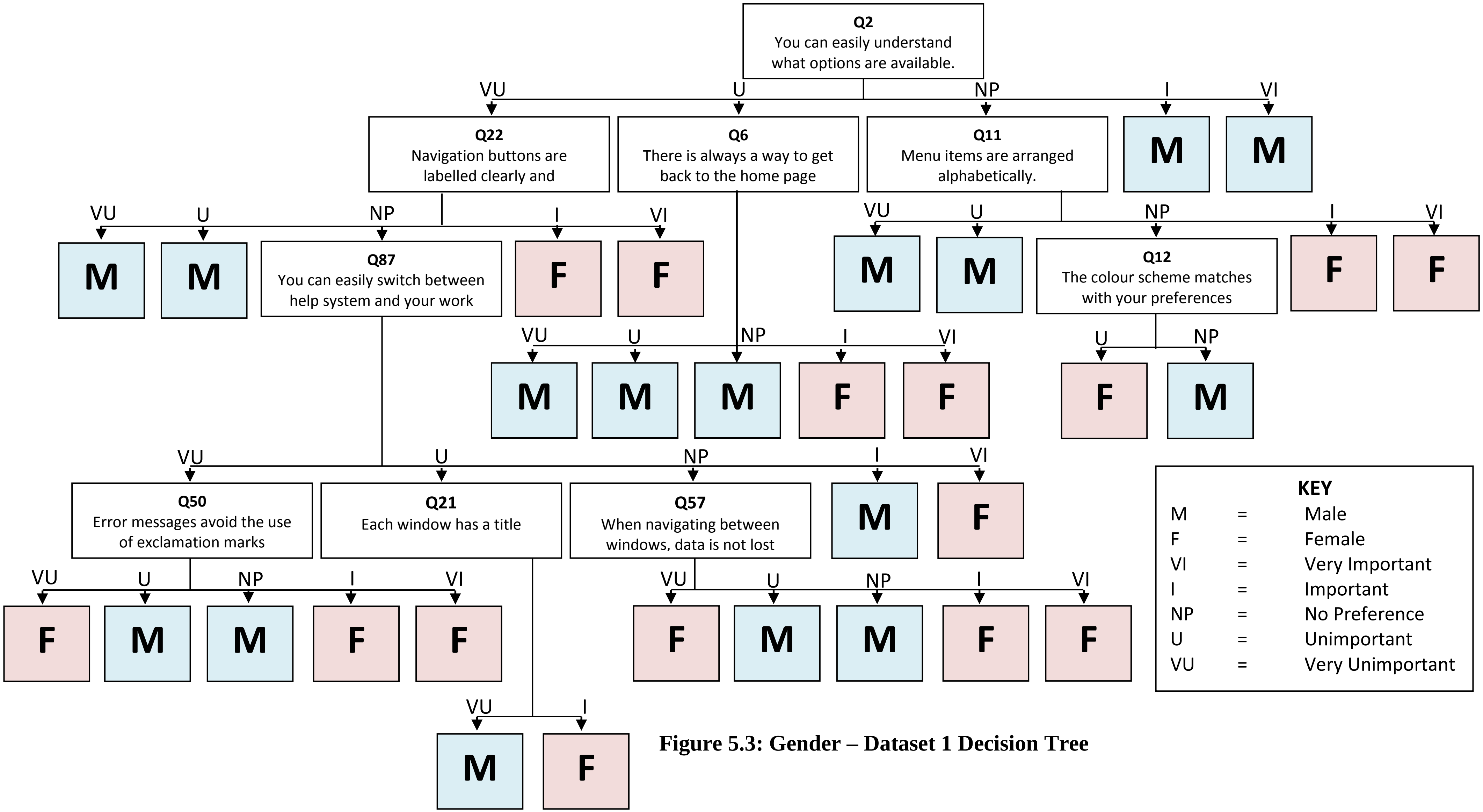



**Figure 5.3: Gender – Dataset 1 Decision Tree**

The decision tree (Figure 5.3) and resulting decision rules (Table 5.2) highlight 9 features. In general, the results show that females need more support from the interface in terms of navigation. For example, females think it important that there was always a way to get back to the homepage, that navigational buttons were labelled clearly and distinctly, that each window has a title and that work is not lost whilst navigating between multiple windows. In contrast, males had no preference or thought these things unimportant. A reason for this could be that in comparison to males, female users tend to experience more difficulty in finding their way effectively around the Internet and are more likely to feel lost and out of control (Ford, Miller and Moss, 2001). The features highlighted by this set of results relate to the way a user can locate information. Therefore, designers can help female users to complete their tasks more efficiently and regain control by providing such elements that will help them to locate information more clearly. For example, Q11 shows that female users think it is important or very important to arrange menu items alphabetically, whilst male users think it unimportant. In addition, it is also shown that males generally prefer to look at the system broadly, as they think it very important that they are easily able to understand what options are available to them (Q2). Females, on the other hand, prefer a much more detailed and linear approach (Roy, Taylor and Chi, 2003). Providing two types of navigation – one linear and one non-linear – that match with male and female navigation styles will help the two groups of users to navigate more efficiently.

**Table 5.2. Decision Rules for Gender Decision Tree**

| Dataset | Gender | Decision Rules |
|---|---|---|
| Dataset 1: Web Searching | M | If users think it important or very important to understand what options are available to them, they are male. |
| | M | If users think it very unimportant to understand what options are available to them, and think it very unimportant or unimportant for navigational buttons to be labelled clearly and distinctively, they are male. |
| | F | If users think it very unimportant to understand what options are available to them, and think it very important or important for navigational buttons to be labelled clearly and distinctively, they are female. |
| | F | If users think it very unimportant to understand what options are available to them, and have no preference for navigational buttons being labelled clearly and distinctively |

### 5.3.3 Statistical Significance

As with the previous two studies, the results were compared with a T-Test. All of the 90 features were included and Table 5.3 shows the 25 found to be statistically significant ($p=<0.05$). With the exception of Q57 and Q87, all of the other features highlighted on the decision tree can be found in this list and thus are statistically significant.

**Table 5.3. Statistical Significance of Features for Gender**

| **Dataset 1: Statistically Significant Features** |
|---|
| Q2, Q3, Q4, Q5, Q6, Q11, Q12, Q13, Q20, Q21, Q22, Q23, Q27, Q37, Q39, Q40, Q45, Q50, Q55, Q56, Q58, Q67, Q68, Q77, Q80 |

## 5.4 Results: Dataset 2 – Web-Based Instruction Tools

### 5.4.1 Feature Selection

As discussed in the previous section, feature selection was performed on the twenty features obtained from Dataset 2 to create two subsets of features. These two subsets were then used to perform decision tree classification and the subset that had the highest classification accuracy was used to create a decision tree to represent users' preferences for WBI tools.

Of the original twenty features, the BN family of classifiers selected 7 and the KNN family selected 8 (Table 5.4). When these two sets are compared, 7 features were commonly selected by both classifiers:

- Q2 ("*Examples given in this tutorial are not practical.*")
- Q6 ("*I would have found it more helpful to be given a suggested route through this tutorial.*")
- Q9 ("*It is hard to use back/forward buttons.*")
- Q10 ("*The information provided by the map is too superficial.*")
- Q11 ("*The links provided in this tutorial help me discover relationships between different topics.*")
- Q12 ("*The map in this tutorial gives a meaningful framework of HTML.*")
- Q13 ("*I was confused which options I wanted, as it provided too many choices.*")

**Table 5.4: Dataset 2: Number of Features Selected per Family**

| | Feature Sets | | | |
|---|---|---|---|---|
| | **Classifier Family** | **Classifier** | **# of Features Selected** | **# of Features in Final Set** |
| **Gender** | BN | BNC | 15 | 7 |
| | | NB | 13 | |
| | | AODE | 11 | |
| | KNN | NN | 17 | 8 |
| | | KNN | 15 | |
| | | K* | 10 | |

### 5.4.2 Decision Tree Classification

Once these two feature sets were identified, three different algorithms were used to build decision trees. Their classification accuracies were then calculated. Like Dataset 1, The C4.5 algorithm had the best overall average and performed the most accurately using the BN feature set. Therefore, this classifier family/decision tree algorithm combination was used to build the decision tree that is used to represent users' preferences for Study 2 (Figure 5.4).

The decision tree for Study 2 highlighted four features: Q10 ("*The information provided by the map is too superficial*"), Q2 ("*Examples given in this tutorial are not practical*"), Q11 ("*The links in this tutorial help to discover the relationships between topics*") and Q19 ("*The tutorial can be used sufficiently well without any instructions*"). Those users that disagreed or strongly disagreed with Q10 were males, whilst those users that agreed or strongly agreed were females. Males have been shown to have less interest in reading Web page content and text than females (Leong and Hawamdeh, 1999), which could perhaps explain why male users are satisfied with the amount of information provided by the map. On the other hand, female users need more detailed information from the map and thus are dissatisfied if the map only contains an overview. Additionally, females have been shown to follow a much more linear pattern than males (Liu and Huang, 2008), whereas males prefer to use the hypertext links within the content (Roy, Taylor and Chi, 2003). This is

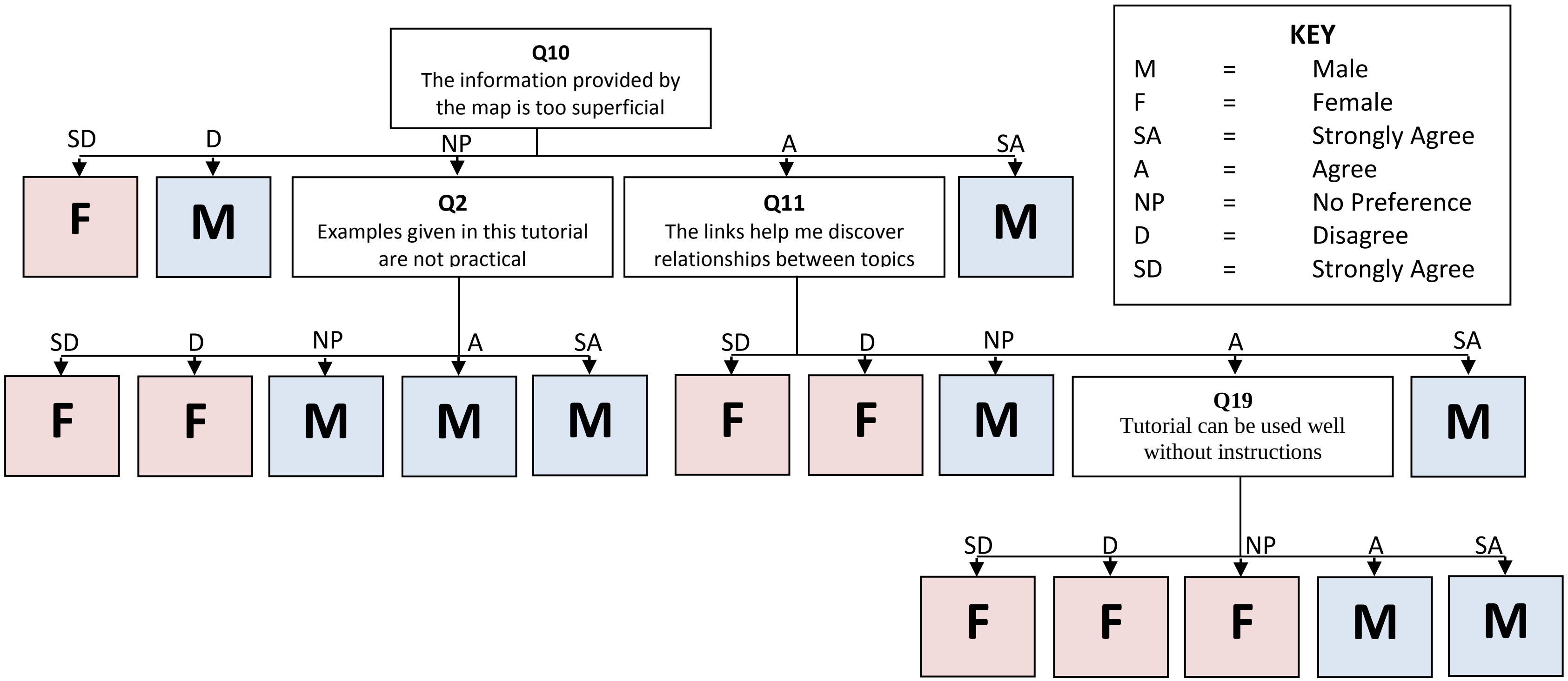



**Figure 5.4: Gender - Dataset 2 Decision Tree**

reinforced by the results of Q11 where users who agreed that the links in the tutorial helped them to discover the relationships within topics were male, whilst those that disagreed were female. In addition, female users disagreed that the tutorial could be used sufficiently well without any instructions (Q19), whereas male users were content with it as it was given. This could be because females tend, in general, to experience more difficulty in finding information online and feel less comfortable when using the Internet (Schumacher and Morahan-Martin, 2001). Furthermore, those users that agreed that the examples in the tutorial were not practical (Q2) were male, whilst those that disagreed were female.

These results for Study 2 show that, in general, males are more positive toward this environment than females. Males tend to use the map to find relationships among topics and think that the map provides a sufficient amount of information. Females, on the other hand, need more additional human support and when interacting with this online environment. They also think that the map should provide more detailed information.

### 5.4.3 Statistical Significance

As with dataset 1, the results of dataset 2 were compared with a T-Test. All of the original 20 features were included and Table 5.5 shows the four found to be statistically significant (p=<0.05) to gender. Two out of the four features highlighted in the decision tree were found to be statistically significant: Q2 and Q11.

**Table 5.5. Statistical Significance for Gender**

| Dataset 2: Statistically Significant Features |
|---|
| Q2 |
| Q6 |
| Q11 |
| Q13 |

## 5.5 Dataset 1 vs. Dataset 2

If we compare the results from these two studies, it is clear that male users place an emphasis on getting a global picture first, whereas female users focus more on depth and getting detailed information. Males prefer to see all the information available and use the hypertext links to begin to filter out the most relevant information at an early stage. On the other hand, females prefer to follow a linear structure, e.g. the map, and build up the relationships between topics step-by-step. These findings agree with previous studies, such as Large, *et al*. (2002), which found that males tend to navigate in a much broader approach than females. Possible reasons for this include males being more familiar, comfortable and actively engaging in the Web environment than females. Designers should take this into consideration when designing the structure of a Web-based application and providing support for these two navigational styles.

In addition, the results also showed that females need more additional support than males in using the Web-based applications. Because females are more uncertain and can easily get lost whilst navigating a website, they will more than likely require additional support in terms of different user interfaces and navigational facilities (Chen and Ford, 1998). This support could be given by providing elements that will help them to locate information more clearly (e.g. clearly labelled buttons or providing a link back to the homepage).

## 5.6 Results: Cognitive Style vs. Previous Experience vs. Gender

As stated in the introduction, the aim of this study is to examine the relationship between three human factors: cognitive style, system experience and gender. The previous section presented the results related to gender, so the following section will begin a detailed analysis of the similarities and differences between the human factors.

### 5.6.1 Dataset 1 – Web Searching

Table 5.6 presents some similarities among these three human factors based on the commonly selected features with BN and KNN classifiers (i.e. features that appear in both the BN and KNN feature sets), which are grouped into three categories, including content organisation, information format and additional support.

On closer inspection of Table 5.6, no one feature was selected by all three human factors. As shown in the fourth and final column in this table, there are the relationships between the human factors. The strongest relationship is between cognitive style and gender, with four common features (i.e., Q9, Q10, Q12, Q21) identified as relevant to both. Interestingly, all of these four features appear in the 'Information Format' category. This suggests that information format is a key factor in the relationship between cognitive style and gender. Among these four features, Q21 ("*Each window has a title.*") is particularly related to both dimensions of cognitive style and to gender.

The second strongest relationship is that of cognitive style and system experience, with three common features (i.e., Q15, Q30, Q77) identified as relevant to both. Apart from Q30, the other two features, i.e. Q15 and Q77, appear in the group of content organisation, which is a key factor for the relationship between cognitive styles and system experience. Among these three features, Q15 ("*Relevant subject content appears on the same page*") is selected as particularly relevant to both dimensions of cognitive style and both dimensions of system experience. On the other hand, the weakest relationship can be seen between system experience and gender, with just one common feature (Q12) identified as relevant to both.

However, although the features in Table 5.6 were selected as relevant to multiple human factors in the feature selection stage, not all of them were highlighted at the classification stage (Table 5.7). Whilst three features were commonly selected by cognitive style and system experience in the feature selection stage, only Q15 is highlighted as a common feature in the classification stage. Furthermore, the results from the classification stage also indicate that Q12 ("*The colour scheme matches with your preferences*") is the only one common feature that links system experience and gender. Moreover, this stage fails to identify any commonly relevant features between cognitive style and gender and between Internet experience and gender.

**Table 5.6: Dataset 1 - Table of Commonly Selected Features**

| Feature No | Description | Category | Selected by Human Factors | | | | |
|---|---|---|---|---|---|---|---|
| | | | CSA | SPQ | IE | CE | Gender |
| Q2 | You can easily understand what options are available to you. | Content Organisation | | | | x | x |
| Q9 | The names of links and buttons are understandable. | Information Format | | | | x | x |
| Q10 | High contrast colour scheme is applied to present the text and background. | Information Format | | x | | | x |
| Q12 | The colour scheme matches with your preferences. | Information Format | | x | | | x |
| Q15 | Relevant subject content appears on the same page. | Content Organisation | x | | | | x |
| Q21 | Each window has a title. | Information Format | x | x | x | x | |
| Q30 | Online instructions appear in a consistent location across screens. | Additional Support | x | x | | | x |
| Q77 | Only information essential to the task is displayed on the screen. | Content Organisation | x | x | | x | |

**Table 5.7. Dataset 1 - Classification Results**

| Feature No | Description | Category | Selected by Human Factors | | | | |
|---|---|---|---|---|---|---|---|
| | | | CSA | SPQ | IE | CE | Gender |
| Q1 | Every page is clearly marked with a label that indicates subject contents. | Information Format | | x | | | |
| Q3 | There is some form of instruction to teach you how to do each activity. | Additional Support | | | | x | |
| Q9 | The names of links and buttons are understandable. | Information Format | | x | | | |
| Q12 | The colour scheme | Information | | | | x | x |

| Feature | Description | Category | Selected by Human Factors | | | | |
|---|---|---|---|---|---|---|---|
| | matches with your preferences. | Format | | | | | |
| Q15 | Relevant subject content appears on the same page. | Content Organisation | x | x | x | x | |
| Q16 | The results are presented by the levels of the relevance. | Content Organisation | x | | | x | |
| Q21 | Each window has a title. | Information Format | | x | | | x |
| Q29 | Error messages appear in the same location throughout the system. | Additional Support | | | | x | |
| Q30 | Online instructions appear in a consistent location across screens. | Additional Support | | | | x | |
| Q31 | Icons are clearly labelled. | Additional Support | | | | x | |
| Q33 | Heavy use of all uppercase letters on a webpage has been avoided. | Information Format | | | | x | |
| Q36 | Messages are appropriate, inoffensive, non-hostile or violent | Additional Support | x | | | | |
| Q37 | When the program presents numbers, integers are right-justified and real numbers are decimal-aligned. | Information Format | | x | | | |
| Q39 | No more than four colours are utilised in each screen. | Information Format | | | | x | |
| Q43 | When you are in trouble, you can receive feedback immediately. | Additional Support | | x | | | |
| Q59 | Data inputs are case-blind whenever possible. | Information Format | x | | | | |
| Q65 | A map that shows the structure of the subject content is available. | Content Organisation | x | x | | | |
| Q75 | The search engine provides you with multiple search options. | Content Organisation | x | x | | | |
| Q78 | The title of each page is short, simple, clear, and | Information Format | | x | x | | |

| Feature | Description | Category | Selected by Human Factors | | | | |
|---|---|---|---|---|---|---|---|
| | distinctive. | | | | | | |
| Q2 | You can easily understand what options are available to you. | Content Organisation | | | | | x |
| Q6 | There is always a way to get back to the home page from any other pages. | Content Organisation | | | | | x |
| Q11 | Menu items are arranged alphabetically. | Content Organisation | | | | | x |
| Q22 | Navigation buttons are labelled clearly and distinctively. | Information Format | | | | | x |
| Q50 | Error messages avoid the use of exclamation points. | Content Organisation | | | | | x |
| Q87 | You can easily switch between the help system and your work. | Content Organisation | | | | | x |

### 5.6.2 Dataset 2 – Web-Based Instruction

As with the first dataset, Table 5.8 presents the similarities among the three human factors based on the commonly selected features with BN and KNN classifiers (i.e. features that appear in both the BN and KNN feature sets). Again, like the first dataset, they can be grouped into three categories, which include content organisation, information format and additional support. Additionally, it can be seen that like the first dataset, no one feature was selected as relevant by all three human factors. However, there are some relationships between the human factors. The strongest relationship can be seen between cognitive style and gender, as three common features (Q2, Q9, Q11) were identified as relevant to both human factors. In particular, Q9 ("*It is hard to use back/forward buttons.*") is related to both dimensions of cognitive style and to gender. The second strongest relationship is that of cognitive style and system experience, with two common features (Q14, Q15). In particular, Q15 ("*I found it hard to select relevant information using the map.*") is selected as relevant to both dimensions of cognitive style and both dimensions of system experience. On the other hand, the weakest can be seen between system experience and gender, with only one common feature (Q12 ("*The map in this tutorial gives a meaningful framework of HTML.*") selected as relevant to both.

Table 5.8 presents the features identified at the classification stage. Whilst there were three features selected in the feature selection stage, two of these features (Q2, Q11) were selected for cognitive style and gender in this second phase. Additionally, the two features that were selected in the feature selection stage were not highlighted in the classification stage, although Q7 was identified as being relevant to both cognitive style and system experience. However, no features are selected for system experience and gender.

Although not as clear as in the first dataset, the third and fourth columns in Table 5.9 show that two out of the four features selected as relevant to gender are categorised as ‘Information Format’. One of these features is relevant to cognitive style and gender, whereas the other is relevant to system experience and gender.

**Table 5.8. Dataset 2 - Table of Commonly Selected Features**

| **Feature No** | **Description** | **Category** | **Selected by Human Factors** | | | | |
|---|---|---|---|---|---|---|---|
| | | | CSA | SPQ | IE | CE | Gender |
| Q2 | Examples given in this tutorial are not practical. | Information Format | | X | | | X |
| Q9 | It is hard to use back/forward buttons. | Content Organisation | X | X | | | X |
| Q11 | The links provided in this tutorial help me discover relationships between different topics. | Content Organisation | X | | | | X |
| Q12 | The map in this tutorial gives a meaningful framework of HTML. | Information Format | | | X | | X |
| Q14 | After using this system I can easily use my knowledge to design home pages. | Additional Support | X | | X | X | |
| Q15 | I found it hard to select relevant information using the map. | Information Format | X | X | X | X | |

**Table 5.9: Dataset 2 - Classification Results**

| Feature No | Description | Category | Selected by Human Factors | | | | |
|---|---|---|---|---|---|---|---|
| | | | CSA | SPQ | IE | CE | Gender |
| 2 | Examples given in this tutorial are not practical. | Information Format | | X | | | X |
| 6 | I would have found it more helpful to be given a suggested route through this tutorial. | Content Organisation | X | X | | | |
| 7 | I would like to have more examples. | Information Format | | X | X | | |
| 9 | It is hard to use back/forward buttons. | Content Organisation | X | X | | | |
| 10 | The information provided by the map is too superficial. | Information Format | | | | | X |
| 11 | The links provided in this tutorial help me discover relationships between different topics. | Content Organisation | | X | | | X |
| 14 | After using this system I can easily use my knowledge to design home pages. | Additional Support | | | X | | |
| 15 | I found it hard to select relevant information using the map. | Information Format | | | X | X | |
| 18 | It is easy to find a route for a specific task with the index. | Information Format | X | X | | | |
| 19 | This tutorial can be used sufficiently well without any instructions. | Additional Support | | | | | X |

### 5.6.3 Dataset 1 vs. Dataset 2

Figure 5.6 summarises the relationships between the three human factors analysed in this study. It is immediately clear to see that the links between cognitive style and gender and cognitive style and system experience are stronger than the link between system experience and gender. Although the previous sections do not provide any evidence that there is a strong link between the three human factors, this nonetheless shows that there are obviously some similarities between them. In addition, the two datasets also highlighted the finding that the features linking cognitive style and gender are related to information format. This was more prominent in the first dataset, with four out of four of the features belonging to this category, whereas only one out of the three was present in the second dataset. However, gender was still one of the human factors with the other two features. This suggests that information format is a key factor in the relationship between cognitive style and gender and designers should take this in consideration when designing interfaces that accommodate two human factors.

It is also noteworthy to mention that more prominent relationships were found in the feature selection stage, rather than the classification stage. The classification aims to further narrow down the number of features so that only the most relevant features related to that particular human factor are included. As the features sets selected in the feature selection stage are not so refined, this suggests that the differences linking the human factors are more general than the key features that are at the base preferences of each human factor.

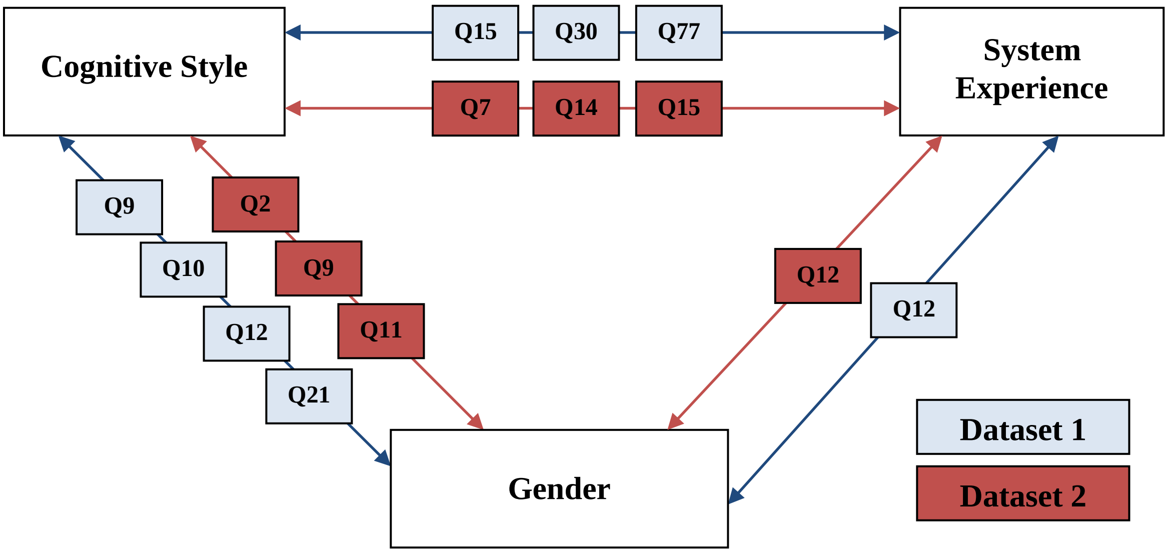



**Figure 5.5: Relationship between Cognitive Style, Gender and System Experience**

# Chapter 6 – Development of an Integrated Model

## 6.1. Introduction

As Web-based applications become accessible to a wider audience, researchers and designers alike are faced with the challenge of identifying the varying effects of human factors and accommodating them in Web-based applications. Among many human factors, cognitive style, previous experience and gender differences significantly influence users' interactions with Web-based applications. Chapter 2 presented a detailed review of the past literature that gives evidence to such influence, whilst Chapters 3, 4 and 5 presented studies that aimed to investigate the relationships between these three human factors upon users' preferences of Web-based applications. More specifically, Chapter 3 investigated the influence of cognitive style and the relationship between two types of cognitive style, Field Dependence/Independence and Holism/Serialism. Chapter 4 investigated the influence of system experience and the relationship between Internet experience and computer experience. Chapter 5 combined the results of the previous two studies and those of the influence of gender differences to investigate the relationship between cognitive style, system experience and gender on users' preferences of Web-based applications.

This chapter begins by presenting an integrated model of the relationships between these three human factors. Based on this model, the chapter then goes on present some design guidelines for the development of more effective Web-based applications.

## 6.2. Development of an Integrated Model

### 6.2.1 Dataset 1: Web Search Tools

Figure 6.1 summarises the findings of these chapters for Dataset 1, Web Searching. Upon closer examination of this figure, it is clear to see that there are four features that are linked by more than one human factor. The following points describe and provide suggestions for the reasons for these relationships.

- Q15 ("*Relevant subject content appears on the same page*") links three dimensions: cognitive style (FD/I), cognitive style (H/S) and Internet experience. Field Independent users, Serialists and users with high levels of Internet experience all find this feature very important. Field Independent users are very analytical and task orientated (Ford *et al.*, 1994) and dislike to spend their time reviewing irrelevant information. In addition, Serialists have a comparatively narrow focus when completing a task and prefer to focus on individual parts of a specific topic (Pask, 1979). Therefore, having relevant content appearing on the same page will help them to maintain this focus. Furthermore, the more Internet experience a user has, the more it is expected that they would be aware of the cues and features that can help them to complete their tasks with more efficiency. Having relevant content that appears on the same page saves the users time in not having to go searching for it.

- Q16 ("*Results are presented by level of relevance*") is linked by two different dimensions: cognitive style (FD/I) and computer experience. Field Independent users find this feature important, whereas Field Dependent users and users with lower levels of computer experience find this unimportant. Field Independent users are able to synthesise important information (Witkin, *et al*., 1977), which they can do quickly and easily if results are presented by level of relevance. On the other hand, Field Dependent users are less analytic and look for the broader picture. They struggle with individual elements and detail, so having results presented to them by level of relevance might not help them attain this global picture. In the same way that Field Dependent users prefer an overall picture, those users with lower levels of computer experience lack the previous experience necessary to gain an understanding of their environment. In this case, they will be more focussed on gaining an overall picture than trying to focus on the links between a specific part.

- Q65 ("*A map that shows the structure of the content is available*") links the two cognitive style dimensions, Field Dependence/Independence and Holism/Serialism. Holists and Field Dependent users find this feature very important, whilst Serialists find it unimportant. Both Holists and Field Dependent users look for a global understanding before seeking out detailed information (Witkin, 1977; Pask, 1979). The map is perfect for helping them to understand the relationships between elements. On the other hand, the Serialist prefers to work through content in detail, topic by topic, only building up the global picture once a global understanding has been reached. An index would be more suitable for a Serialist.

- Q21 ("*Each window has a title*") links two dimensions: cognitive style (H/S) and gender differences. Holists and males find this feature very unimportant, whilst Females find this feature very important. Females tend to get more confused with the surrounding environment and so would look for cues such as window titles or link titles to orientate themselves (Riding and Al-Sanabani, 1998). Holists, on the other hand, are internally motivated and are able to create conceptual links without the need to rely on such cues.

- Q12 ("*The colour scheme matches with your preferences*") links two dimensions: computer experience and gender differences. Both female users and those with lower levels of computer experience find this unimportant. When involved in a task, those users with lower levels of experience would focus more on learning what is needed to complete the task (i.e. which button to click), rather than the aesthetics of the surrounding environment. This is the same for females, who tend to be more anxious in online environments (Abbott and Bienvenue, 2007).

### 6.2.2 Dataset 2: Web-based Instruction Tools

Figure 6.2 summarises the findings of the previous three chapters for Dataset 2, Web-based instruction tools. Upon closer examination of this figure, it is clear to see that there are five features that are linked by more than one human factor. The following points describe and provide suggestions for the reasons for these relationships.

- Q11 ("*The links provided in this tutorial help me to discover the relationships between different topics*") links two dimensions: cognitive style (H/S) and gender differences. Females and Holists disagree with this feature, whilst males and Serialists agree. Holists are internally motivated and are able to compile linking structures by themselves. Serialists, on the other hand, are externally motivated and look to structure cues such as links to navigate content. In addition, females are more exhaustive in their search for information and rely on a broad range of information from a variety of external sources (Meyers-Levy, 1988). Thus, females would require more links than males.

- Q7 ("*I would have liked to have more examples*") links two dimensions: cognitive style (H/S) and Internet experience. Serialists tend to analyse theory and application separately, concentrating first on understanding (Pask, 1979). In addition, higher levels of Internet experience will already have the application knowledge and this will focus more on theoretical understanding. Lower levels of Internet experience, however, will not have this application knowledge and will therefore need a base (i.e. examples) with which to understand the theoretical content.

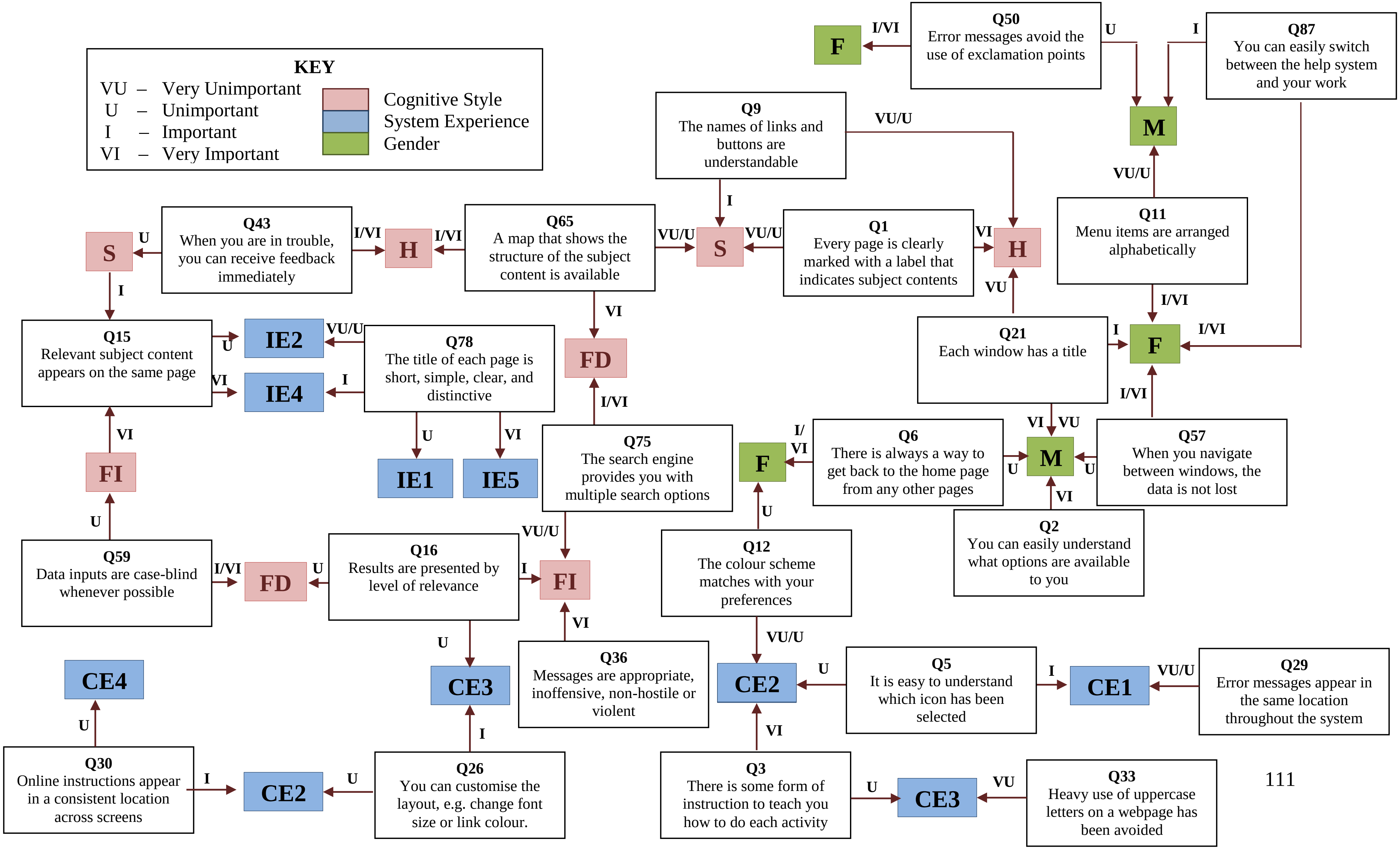
- Q2 ("*Examples in this tutorial are not practical*") links two dimensions: cognitive style (H/S) and gender differences. Males, Serialists and Holists agree with this feature, whereas females strongly disagree. This finding suggests that Holists and Serialists show similar preferences. However, females and males reacted to this issue differently. More specifically, females appreciate the examples presented in this tutorial but males need highly practical examples.

- Q6 ("*I would have found it more helpful to be given a suggested route though this tutorial*") links three dimensions: cognitive style (FI/FD), cognitive style (H/S) and gender differences. Females strongly disagree with this feature, whereas Field Dependent users and Serialists strongly agree with this feature. Field Dependent users and Serialists are externally motivated (Witkin, 1977;

Pask, 1979) and will therefore prefer cues and suggested content structure that will help them to navigate the content. Females prefer a more exhaustive search and draw information in from a variety of external references (Meyers-Levy, 1988). Perhaps, therefore, a suggested route is not suitable for them because it is not a comprehensive enough approach.

- Q18 ("*It is easy to find a route for a specific task in the index*") links the two cognitive style dimensions: Field Dependence/Independence and Holism/Serialism. Holists disagree with this feature, whereas Serialists and Field Independent users agree with this feature. Serialists and Field Independent users prefer a local analytical approach to approaching a task (Pask, 1979; Witkin, 1977). Therefore, an index, which provides a detailed listing often in alphabetical order, is perfect for them as they are not concerned with the global understanding until a local understanding has been reached. Holists, on the other hand, prefer to get a global picture at the beginning and an index does not provide this facility.

**Figure 6.1. Integrated Model of the Influence of Human Factors on Users' Preferences of Web Search Tools**



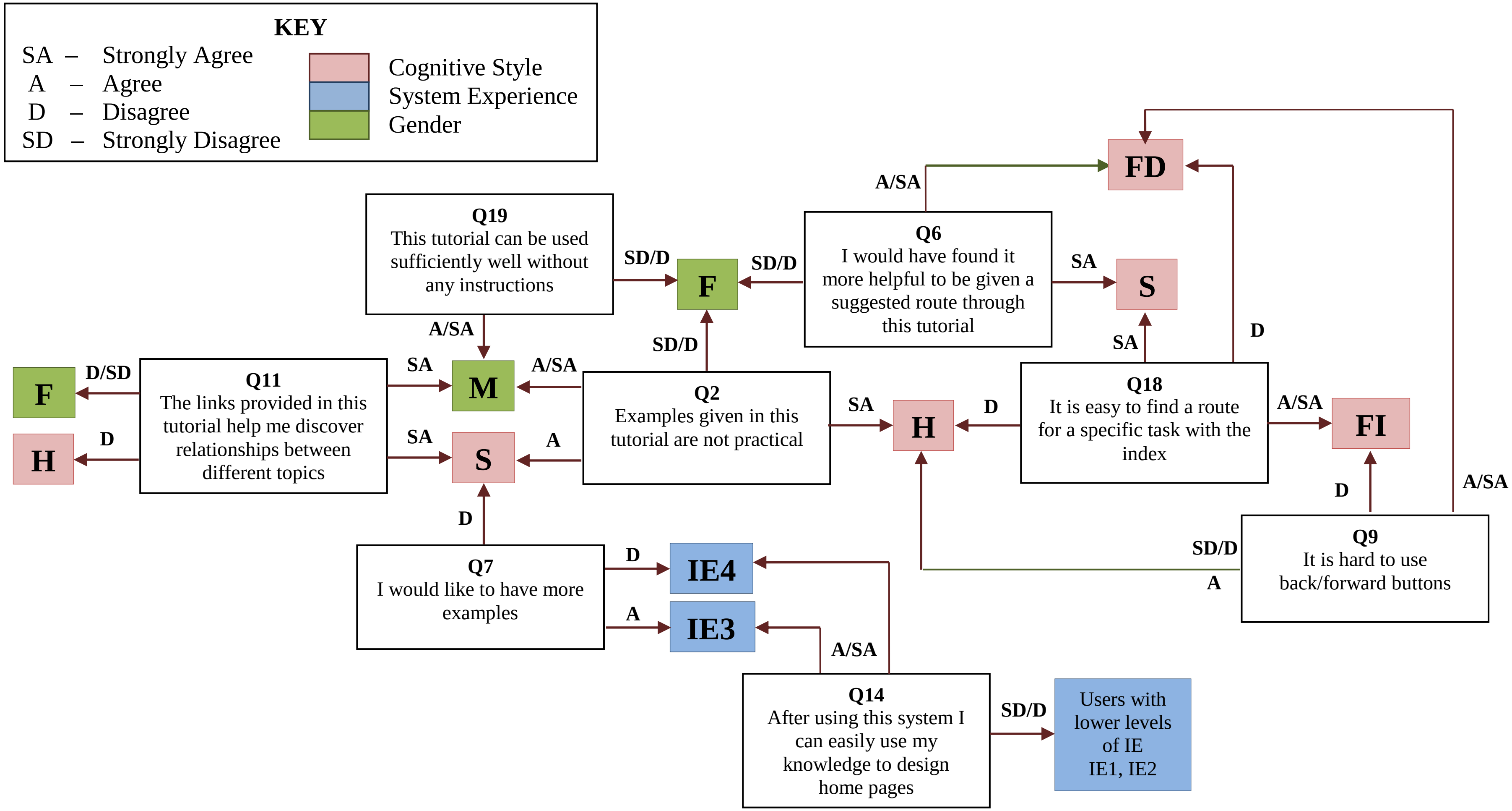



**Figure 6.2. Integrated Model of the Influence of Human Factors on Users' Preferences of Web-Based Learning Tools**

## 6.3. Implications for System Design

The integrated models shown in Figures 1 and 2 have implications for the design of Web-based applications. Table 6.1 presents a summary of the key elements taken from the models that designers should take into consideration when developing effective Web-based applications.

Upon closer analysis of Table 6.1, it is clear to see that specific categories of guidelines are more relevant to some human factors than to others. For example, the categories of navigation and content presentation seem to be very important to cognitive styles. These categories refer to the way users organise and represent information, which is significant because this is the definition of cognitive style (Riding and Rayner, 1998). Thus, designers should pay specific attention to these categories when designing for users with different cognitive styles. In addition, the category of Information Format seems to be especially important to system experience. The higher a user's level of experience, the more they will be aware of which features will help them to complete their tasks with more efficiency. Therefore, having information in a format that allows them to locate it efficiently will help them to achieve this task. Designers who are developing Web-based applications for users with different levels of system experience should take particular notice of these issues. Furthermore, content presentation and additional support seem to be particularly important to gender differences.

## 6.4. Benefactors

The goal of this study was to investigate the influence of human factors on users' preferences so that a framework could be developed that can help to inform the design of Web-based applications. This framework was developed in Section 6.2. Such a framework is beneficial to several parties.

- **System Designers**

  First and foremost, the framework will be a benefit for system designers of Web-based applications. The design of usable and effective Web-based applications relies upon the design being compatible with users' characteristics (Dillon and Zhu, 1997). This investigation looked at how users' characteristics (human factors) affected their preferences and this is captured within the framework. Therefore, designers are empowered by applying this framework and considering the design implications.

- **Users/Students**

Directly as a result of allowing system designers to create more usable and effective systems, the framework will bring long term benefits to the end users and students of the Web-based applications. Interacting in an environment that more effectively supports their needs and preferences will help to promote their learning and task performance.

- **Academics and Researchers**

This investigation will be of interest to many different academic disciplines, including data mining, human-computer interaction (HCI), and information retrieval. In particular, the framework is of great interest because it is one of the very few studies that take a step in the direction of analysing the combined effects of multiple human factors.

- **Other types of Web-based applications**

Although the framework was developed using only two Web-based applications, the framework can be generalised across a multitude of applications, for example, Internet shopping applications. This would help to enhance the effectiveness of all Web-based applications by helping designers to better understand users’ needs and preferences.

**Table 6.1. Design Implications for Web-Based Applications**

| Design Guidelines | Cognitive Style | | | | System Experience | | | | Gender | |
|---|---|---|---|---|---|---|---|---|---|---|
| | FD | FI | Holist | Serialist | IE Novice | IE Expert | CE Novice | CE Expert | Male | Female |
| **Navigation** | | | | | | | | | | |
| Index | | X | | X | | | | | | |
| Hierarchical Map | X | | X | | | | | | | |
| Main menu is alphabetically ordered | | | | | | | | | | X |
| Back/forward buttons | | X | | | | | | | | |
| **Content Presentation** | | | | | | | | | | |
| Suggested route through content | X | | | X | | | | | | |
| ‘Home’ link on every page | | | | | | | | | | X |
| Multiple options for completing the same tasks | X | | | | | | | | | |
| Different options are clearly shown | | | | | | | | | X | |
| Provide links within the text that show the relationships between subjects | | | | X | | | | | X | |
| Relevant subject content appears on the same page | | X | | | | X | | | | |
| Results presented by level of relevance | | X | | | | | | | | |
| **Information Format** | | | | | | | | | | |
| Case-blind data inputs | X | | | | | | | | | |
| Each window has a title that is short, distinctive and relevant to subject content | | | X | | | X | | | X | X |
| User can customise layout to their preferences (i.e. font style, size and colour) | | | | | | | | X | | |
| Easy to understand what icons are selected. | | | | | | | X | | | |
| **Additional Support** | | | | | | | | | | |
| Lots of practical examples | | | X | X | X | | | | X | X |
| Immediate feedback | | | | | | | | | | |
| System messages and instructions appear in consistent locations across screens | | | | | | | X | | | |
| Ability to switch between help system and work with data not being lost | | | | | | | | | X | X |

| Design Guidelines | Cognitive Style | | | | System Experience | | | | Gender | |
|---|---|---|---|---|---|---|---|---|---|---|
| | FD | FI | Holist | Serialist | IE Novice | IE Expert | CE Novice | CE Expert | Male | Female |
| Additional support for each action is provided in the form of instructions or help. | | | | | | | | | | X |
| Avoid use of ! in error messages | | | | | | | | | | X |
| Messages are appropriate, inoffensive, non-hostile and non-violent. | | X | | | | | | | | |
| Provide feedback immediately if user gets into trouble | | | X | | | | | | | |
| Additional support needed after learning to apply knowledge in another scenario | | | | | | | | | | X |
| Additional instructions about how to complete each task is given. | | | | | | | X | | | |

# Chapter 7 – Conclusion

## 7.1 Introduction

The aim of this chapter is to briefly review the process and outcomes of the investigation reported within this thesis. In the very first chapter of this thesis, two aims and three research questions were identified. The achievement of these aims is reviewed in the following sections before the limitations and future research directions of this study are discussed.

## 7.2 Investigating the Relationships Between Human Factors

The first of the main aims of this thesis was to investigate the relationships between three human factors – cognitive style, system experience and gender differences – using users' preferences of Web-based applications. As Web-based applications become more widespread, designers have to accommodate the needs of a wider audience with a diverse set of preferences. Investigating the human factors that affect users' preferences will help both researchers and designers to develop effective Web-based applications for the future. In recognising the importance of such an investigation, a detailed review of existing literature was conducted in Chapter 2. From this, Chapter 3, 4 and 5 conducted three studies that aimed to provide answers to this aim through answering three key research questions. These research questions are stated in the following sections with the key findings from the corresponding study.

### 7.2.1 Research Question 1

*In terms of cognitive style, to what extent is the similarity between the two dimensions Witkin's Field Dependence/Independence and Pask's Holism/Serialism?*

Chapter 3 presented Study 1, which investigated the influence of cognitive style on users' preferences for Web-based applications. More specifically, the relationship between the two cognitive style dimensions, Field Dependent/Independent and Holist/Serialist, was analysed with the aim of identifying whether previous researchers are correct in their assumptions that Field Dependent individuals have similar preferences to Holists, and Field Independent individuals have similar preferences to Serialists. The results of the study suggested that whilst these two relationships were the strongest in both datasets, there were relationships between Field Dependent individuals and Serialists, and Field Independent individuals and Holists. This contradiction appears in the results for Dataset 2, for the feature Q6 ("*I would have found it more helpful to be given a suggested route through the tutorial*"). Field Independent users and Holists were shown to disagree with this statement, whilst Field Dependent users and Serialists were shown to agree with this statement. Possible reasons for this might be because it is in the nature of both Field Dependent users and Serialists to be externally directed and prefer a structured presentation of information (Witkin *et al.*, 1977; Pask, 1979). On the other hand, Field Independent users and Holists are more internally directed and are better able to impose their knowledge on the structures around them. Thus, they are able to flourish in an environment that allows them to choose their own paths.

### 7.2.2 Research Question 2

*In terms of system experience, to what extent do computer experts have the same preferences as Internet experts and computer novices have the same preferences as Internet novices?*

Chapter 4 presented Study 2, which investigated the influence of system experience on users' preferences for Web-based applications. In particular, the relationship between the two dimensions of system experience, Internet experience and computer experience, was analysed. The aim was to confirm or disprove the hypothesis that Internet novices are like computer novices and Internet experts are like computer experts. The results showed that although there are some similarities between the two dimensions, the two are not equally alike. More specifically, this is shown by the results of Dataset 2, through the appearance of feature Q15 ("*I found it hard to select relevant information using the map*"). The more Internet experience users had, the easier it was for them to find relevant information using the map structure. However, this did not necessarily mean that the more computer experience students had, the easier it was for them to find information using the map structure, as those with high levels of computer experience found it difficult to use the map structure to find relevant information. This suggests that computer experience is not entirely comparable to Internet experience, as users would have difficulty in applying their computer knowledge to an Internet environment.

### 7.2.3 Research Question 3

*In terms of the relationship between the three human factors, to what extent are Field Independent users, experts and males alike, and Field Dependent users, novices and females alike?*

Chapter 5 presented Study 3, which investigated the influence of gender on users' preferences for Web-based applications. In addition, the study in this chapter combined these gender results with the results from Study 1 and Study 2 to analyse the relationship between the three human factors: cognitive style, system experience and gender. The aim of this investigation was to confirm or disprove the hypothesis that Field Independent users, experts and males are alike and Field Dependent users, novices and females are alike. Although this study could not provide any definitive evidence for such a relationship, there were nonetheless relationships defined among the human factors. In particular, the relationships between cognitive style and gender and between cognitive style and system experience were found to be stronger than the relationship between system experience and gender. Furthermore, the features that link cognitive style and gender were found to be related to information format, which suggests this is a very important factor in the relationship between the two.

## 7.3 Identifying a Suitable Analysis Method

The second main aim of this thesis was to identify a suitable data analysis methodology that could a) overcome the natural fuzziness of human preferences data, b) avoid the need to make predefined assumptions about the data, and c) have the ability to compare multiple human factors at the same time.

This aim was achieved by using an integrated data analysis model that combined the data mining techniques of feature selection and decision tree classification. This method was successful in terms of overcoming the natural fuzziness of human preferences data as the feature selection stage help to 'weed out' the irrelevant data. In addition, the classification stage enabled the identification of the most accurately classified feature set. Thus, the most accurate feature set was classified with the most accurate decision tree algorithm to produce what we can assume to be the decision tree that most accurately demonstrates users' preferences for both Web search engines and Web-based instruction tools.

Although this method proved very effective when analysing the relationships between different dimensions of the same human factor (i.e. Field Dependence/Independence and Holism/Serialism, or Internet experience and computer experience), the results were not so fruitful when analysing the relationships between the human factors themselves (i.e. cognitive style, system experience and gender differences). There could be two possible reasons for this. Firstly, there truly is no relationship between these three human factors, at least within the datasets studied within this thesis. Further investigations involving multiple other datasets could confirm and test this issue. Secondly, the integrated data mining data analysis model was used in order to create highly relevant decision trees for a specific dimension of a human factor. Thus, only the most relevant issues for each dimension are highlighted within the decision tree. Perhaps this method is too specific, especially if the characteristics that link the human factors are not included in these most relevant issues, rather they are located slightly farther afield. This hypothesis is in agreement with the results of Study 5, as the majority of features that linked the human factors were found in the feature selection stage, i.e. before only the most relevant features were selected. This issue can be tested through performing additional investigations where both the feature selection results and the decision tree classification results are compared and combined.

## 7.4 Significance of Thesis

The significance of this thesis lies within two different aspects, including theory and methodology.

- In regards to theory, this investigation helps to deepen the understanding of the differences of different types of users within different dimensions of human factors, i.e. Field Dependent users, Field Independent users, Holists, Serialists, Internet novices, Internet experts, computer novices, computer experts, males and females. In addition, this investigation also helps to identify the interactions between human factors, taking a step forward into understanding the combined effects of multiple human factors. Suggestions for design and development were proposed to assist in the continued effectiveness of Web-based applications.

- In regards to methodology, this study uses an integrated data mining analysis approach that combines the techniques of feature selection and decision tree classification. The main significance of this methodology lies within its ability to accurately classify the most relevant feature sets in order to identify the decision trees that most accurately demonstrates users' preferences for Web-based applications.

### 7.5 Limitations and Suggestions for Future Work

Every piece of research has its associated limitations that can undoubtedly affect the conclusions reached. However, these limitations can often provide a starting point for future research. The following sections summarise the limitations of the research presented in this thesis and suggest possible avenues for future research.

- The investigation in this thesis was carried out using just two datasets, one that contained users' preferences for Web-searching and one that contained users' preferences for Web-based instruction tools. Although it was possible to identify a set of design guidelines based upon the analysis of these two datasets, future research should expand the investigation into other Web-based applications. For example, Internet banking or e-Stores. By doing so, the findings of this investigation can be verified and strengthened. In addition, it would be interesting to see if the results of this study are common across multiple Web-based applications.

- In Study 1, users' cognitive styles were identified using just two tests: the Cognitive Styles Analysis (CSA) to determine Field Dependence/Independence and the Study Preference Questionnaire (SPQ) to determine Holism/Serialism. Further works should look at using additional tests to identify users' cognitive styles, such as the Group Embedded Figures Test (GEFT), to see if the same results are obtained.

- In Study 2, users' level of Internet and computer experience was measured by the users themselves. Consequently, users may not have an accurate idea of their own levels of experience and thus this self-rating could influence the accuracy of the results. Further research could use other more objective measures, e.g. quantitative measures of previous exposure to computers and the internet, to verify these results presented in this study.

- In Study 3, the relationship between just three human factors were analysed, including cognitive style, system experience and gender differences. There are a multitude of other human factors, such as age and domain knowledge, so it would be interesting to identify the relationships between these human factors. In addition, it would be interesting to see if the relationship between specific human factors is highlighted in specific Web-based applications. The benefits of doing this would be to reduce the amount of different interfaces and instead focus on the more significant factors to help develop effect Web-based applications.

- In terms of the integrated data mining approach, only six classifiers from two different families were used in the feature selection stage to select relevant features. In addition, only three algorithms were used in the decision tree classification stage. It would be interesting to extend the range of classifiers (e.g. SVM or Neural Networks) and decision tree algorithms (e.g. CHAID) to see if similar results are obtained.

- As previously mentioned, this thesis investigated the relationships between human factors through the use of two datasets. These datasets were small in size, i.e. Dataset 1 contained 90 features with 120 data instances and Dataset 2 contained 20 features with only 65 data instances. In the original application areas of data mining techniques, such as Bioinformatics, thousands of features and data instances are used. Although getting hold of a user preferences dataset this size would be very difficult, it would be interesting to see if the integrated data mining analysis approach would produce the same results.

# References


Abbott, G., & Bievenue, L. (2007). Gender equity in the use of Educational technology. In Klein (Ed.). *Handbook for achieving gender equity through education*, pp. 191-214. Lawrence Erlbaum Associates.

Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining Associations between Sets of Items in Large Databases. *ACM SIGMOD Int'l Conf. on Management of Data*, Washington D.C., pp. 207-216.

Ankerst, M. (2001). Human Involvement and Interactivity of the Next Generation's Data Mining Tools. *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, Santa Barbara, CA, 2001

Ash, B. (1986). Identifying learning styles and matching strategies for teaching and learning. Massachusetts: ERIC Document Reproduction Service (Education Resources Information Center, ED 270142).

Baeza-Yates, R. and Raghavan, P. (2010). Next Generation Web Search. In Ceri, S. and Brambilla, M. (Eds): *Search Computing*, LNCS 5950, Springer-Verlag Berlin Heidelberg, pp. 11-23.

Baglioni, M., Furletti, B., Turini, F. (2005). Means of prior knowledge. In *ACM symposium on Applied Computing,* pp. 474–481.

Bar-Ilan, J. (2005). Information hub blogs. *Journal of Information Science*, 31, 4, pp. 297-307.

Bohen, S.P., Troyanskaya, O.G., Alter, O., Warnke, R., Botstein, D., Brown, P.O., Levy, R. (2003). Variation in gene expression patterns in follicular lymphoma and the response to rituximab. In *Proceedings of the National Academy of Sciences of the United States of America*, 100, 4, pp. 1926-1930.

Bradlow, T., Hoch, S.J., and Hutchinson, J.W. (2002). An assessment of basic computer proficiency among active Internet Users: test construction, calibration, antecedents, and consequences. *Journal of Educational and Behavioral Statistics*, 27, 1, pp. 237-253.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., (1984). *Classification and Regression Trees*. Wadsworth Int. Group/Probability Series, Belmont, California, USA.

Brumby, M.N. (1982). Consistent differences in cognitive styles shown for qualitative biological problem-solving. *British Journal of Educational Psychology*, 52, 244–257.

Castañeda, J.A., Muñoz-Leiva, F., and Luque, T. (2007). Web Acceptance Model (WAM): Moderating effects of user experience. *Information & Management*, 44, 4, pp. 384-396.

Chang, L. Y., and Chen, W. C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36, pp. 365–375.

Chen, S. Y. (2005). Evaluating the Learning Effectiveness of using Web-based Instruction: An Individual Differences Approach. *International Journal of Information & Communication Technology Education*, 1, 1, pp. 69-82.

Chen, S. Y. (2000). *The Role of Individual Differences and Levels of Learner Control in Hypermedia Learning Environments*. Ph.D. Thesis, University of Sheffield, UK.

Chen, S. Y., Fan, J and Macredie, R. D. (2006). Navigation in Hypermedia Learning Systems: Experts vs. Novices. *Computers in Human Behavior*, Computers in Human Behavior, 22, 2, pp. 251-266.

Chen S.Y. & Ford N. (1998). Modelling user navigation behaviours in a hypermedia-based learning system: an individual differences approach. *International Journal of Knowledge Organization*, 25, pp. 67–78.

Chen, Y., Hsu, C. and Chou, S. (2003). Constructing a multi-valued and multi-labeled decision tree. *Expert Systems with Applications*, 25, 2, pp. 199-209.

Chen, S. Y. and Liu, X. (2008). An Integrated Approach for Modeling Learning Patterns of Students in Web-based instruction: A Cognitive Style Perspective. *ACM Transactions on Computer-Human Interaction*, 15, 1.

Chen, S. Y. and Macredie, R. D. (2010). Web-based Interaction: A Review of Three Important Human Factors. *International Journal of Information Management*, 30, 5, pp. 379-287.

Chen S.Y. and Macredie R.D. (2004). Cognitive modelling of student learning in web-based instructional programmes. *International Journal of Human-Computer Interaction*, 17, pp. 375–402.

Chen S.Y. & Macredie R.D. (2002). Cognitive styles and hypermedia navigation: development of a learning model. *Journal of the American Society for Information Science and Technology,* 53, pp. 3–15.

Chrysostomou, K.A. (2008) *The Role of Classifiers in Feature Selection: Number vs Nat*ure, PhD Thesis, Brunel University, UK.

Chrysostomou, K.A., Lee, M., Chen, S.Y., and Liu, X., (2008). Wrapper Feature Selection. *Encyclopedia of Data Warehousing and Mining*, 4, pp. 2103-2108.

Chuang, S. and Tsai, C. (2005). Preferences toward the constructivist internet-based learning environments among high school students in Taiwan. *Computers in Human Behavior*, 21, pp. 255-272.

Clark, P. and Niblett, T. (1989). The CN2 induction algorithm. *Machine Learning*, 3, pp. 261-283.

Cutmore, T.R.H, Hine, T.J., Maberly, K.J., Langford, N.M. and Hawgood, G. (2000). Cognitive and gender factors influencing navigation in a virtual environment. *International Journal of Human Computer Studies*, 53, 2, pp. 223–249.

Dee-Lucas, D., & Larkin, J. H. (1995). Learning from electronic texts: Effects of interactive overviews for information access. *Cognition and Instruction*, 13, 3, pp. 431–468.

de Souza, J.T., Matwin, S. and Japkowicz, N. (2006). Parallelizing Feature Selection. *Algorithmica*, 45, 3, pp. 433-456.

DeTure, M. (2004). Cognitive Style and Self-Efficacy: Predicting student success in online distance education, *American Journal of Distance Education*, 18, 1, pp. 21-38.

Dillon, A., and Zhu, E. (1997). Design Web-Based Instruction: A Human-Computer Instruction Perspective. *Web-based instruction* (pp.221-224). Englewood Cliffs, N.J. Educational Technology Publication.

Dufresne, A. and Turcotte, S. (1997). Cognitive style and its implications for navigation strategies in B du Boulay and R Mizoguchi (eds) *Artificial intelligence in education knowledge and media learning system*, Kobe (Japan), Amsterdam IOS Press, pp. 287–293.

Durndell, A. and Haag, Z. (2002). Computer self-efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior,* 18, pp. 521–535.

Ellis, D., Ford, N. and Wood, F. (1992). Hypertext and learning styles. *Final report of a project funded by the Learning Technology Unit*. Sheffield: Employment Department.

Entwistle, N. (1981). Styles of learning and teaching: An integrated outline of educational psychology—For students, teachers and lecturers. Chichester: John Wiley.

Fan, J. (2005) *Interface design for hypermedia learning systems: a study of individual differences and system feature*, PhD Thesis, Brunel University, UK.

Farrell, I.H., & Moore, D.M. (2000). The effects of navigation tools on learners' achievement and attitude in a hypermedia environment. *Journal of Educational Technology Systems*, 29, 2, pp. 169-181.

Flanigan, A., and Metzger, M. (2003). The perceived credibility of personal web page information as influenced by the sex of the source. *Computers in Human Behaviour,* 19, 6, pp.683–701.

Ford, N. (2000). Improving the "darkness to light ratio" in information retrieval research. *Journal of Documentation*, 56, 6, pp. 624–643.

Ford, N. (1985). Learning styles and strategies of postgraduate students. British *Journal of Educational Technology*, 16, 65– 77.

Ford N. and Chen S.Y. (2001). Matching/mismatching revisited: an empirical study of learning and teaching styles. *British Journal of Educational Technology*, 32, pp. 5–22.

Ford N. and Chen S.Y. (2000). Individual differences, hypermedia navigation and learning: an empirical study. *Journal of Educational Multimedia and Hypermedia*, 9, pp. 281–311.

Ford, N., Miller, D. and Moss, N. (2005). Web search strategies and human individual differences: Cognitive and demographic factors, Internet attitudes, and approaches. *Journal of the American Society for Information Science and Technology*, 56, pp. 741-756.

Ford, N., Miller, D. and Moss, N. (2001). The role of individual differences in Internet searching: An empirical study. *Journal of the American Society for Information Science and Technology*. 52, pp. 1049-1066.

Ford, N., & Miller, D. (1996). Gender differences in Internet perceptions and use. *Aslib Proceedings*, 48, pp. 183–192.

Ford, N., Wilson, T.D., Foster, A., Ellis, D., and Spink, A. (2002). Information seeking and mediated searching. Part 4: Cognitive styles and individual differences. *Journal of the American Society for Information Science and Technology*, 53, 9, pp. 728-735.

Ford, N., Wood, F., & Walsh, C. (1994). Cognitive styles and online searching. Online *and CD-ROM Review*, 18, 2, pp. 79–86.

Frias-Martinez, E., Chen, S.Y. Macredie, R, and Liu, X. (2007). The Role of Human Factors in Stereotyping Behavior and Perception of Digital Library Users: A Robust Clustering Approach. *User Modelling and User Adapted Interaction*, 17, 3, pp. 305-337.

Gammerman, A.J. (1997). *Machine learning: progress and prospects*. Egham, Surrey, Royal Holloway, University of London.

Gauss, B. & Urbas, L. (2003). Individual differences in navigation between sharable content objects - an evaluation study of a learning module prototype, *British Journal of Educational Technology*, 34, 4, pp.499-509.

Graff, M. (2003). Learning from web-based instructional systems and cognitive style. *British Journal of Educational Technology*, 34, pp. 407–418.

Gunn, C., M. McSporran, H. Macleod, and S. French. (2003). Dominant or Different? Gender Issues in Computer Supported Learning. *Journal of Asynchronous Learning Networks* 7, 1, pp. 13–30.

Han, J. and Kamber, M., (2006). *Data Mining: Concepts and Techniques*, Second Edition, The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers.

Hand, D.J., Mannila, H., and Smyth, P. (2001). *Principles of data mining*. MIT Press.

Hölscher, C. and Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, 33, pp. 337–346.

Howard, J.S. and Smith, R.D. (1986). Computer anxiety in management: myth or reality? *Communications of the ACM*, 29, 7, pp. 611-615.

Hsu, P., Lai, R. And Chiu, R. (2003). The hybrid of association rule algorithms and genetic algorithms for tree induction: an example of predicting the student course performance. *Expert Systems Applications*, 25, 1, pp. 51-62.

Hudson, L. (1966). *Contrary imaginations: A psychological study of the English schoolboy*. London, Methuen.

Huang, C.H., Liu, M.C., Chu, S.S and Cheng, C.L. (2007). An intelligent learning diagnosis system for Web-based thematic learning platform. *Computers & Education*, 48, 4, pp. 658-679.

Hyde, J.S. (2005). The gender similarities hypothesis. *The American Psychologist*, 60, 6, pp. 581-592.

Ivory, M.Y. and Megraw, R. (2005). Evolution of web site design patterns, *ACM Transactions on Information Systems (TOIS)*, 23, 4, pp. 463-497.

Jackson, L. A., Ervin, K. S., Gardner, P. D. and Schmitt, N. (2001). Gender and the Internet: women communicating and men searching, *Sex Roles*, 44, pp. 363–79.

Jenkins, C., Corritore, C.L. and Wiedenbeck, S. (2003). Patterns of Information Seeking on the Web: A qualitative study of domain expertise and Web expertise, *Information Technology and Society*, 1, 3, pp.64-89.

Jonassen, D. H., & Grabowski, B. (1993). *Individual differences and instruction*. New York: Allyn & Bacon.

Kagan, J. (1966). Reflection and impulsivity: The generality and dynamics of conceptual tempo. *Journal of Abnormal Psychology*, 71, pp. 17-24.

Kass, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 29, 2, pp. 119-127.

Katz, J. E., Rice, R.E. and Aspden, P. (2001). The Internet, 1995–2000. *American Behavioral Scientist*, 45, pp. 405 – 19.

Keim, D.A. (2002). Information visualization and visual data mining, *IEEE Transactions on Visualization and Computer Graphics*, 8, 1, pp. 1-8.

Khan, B.H. (2005). *Managing e-learning: Design, delivery, implementation and evaluation*. Hershey, PA: Information Science Publishing.

Khan, B.H. (1998). Web-Based Instruction (WBI): An Introduction, *Educational Media International*, 35, 2, pp. 63-71.

Kim, S.J., Suh, E.H. and Yoo, K.D. (2007). A study of context inference for Web-based information systems. *Electronic Commerce Research and Applications*, 6, pp. 146–158.

Kim, K.S. (1997). *Effects of cognitive and problem-solving styles on information-seeking behavior in the WWW: A case study* [On-line]. Available: http://www.edb.utexas.edu/mmresearch/Students07/Kim.

Kim, H., Yun, M. and Kim, P. (2004). Comparison of Web Searching Strategies According to Cognitive Styles of Elementary Students. *Lecture Notes in Computer Science*. Vol., 3044, p. 892-901

Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R. and Suh, J.H. (2002). A personalized recommendation procedure for Internet shopping support. *Electronic Commerce Research and Applications*. 1(3/4), pp. 301-313.

Kirton, M. J. (1976) Adaptors and Innovators: A Description of a Measure, In the Journal of Applied Psychology, 61, p. 622-629.

Kominski, R. and Newburger, E. (1999). Access Denied: Changes in Computer Ownership and Use: 1984–1997, In Proceedings of the annual meeting of the American Sociological Association. Chicago, Ill.

Kraut, R., Scherlis, W., Mukhopadhyay, T., Manning, J., & Kiesler, S. (1996). The HomeNet field trial of residential Internet services, *Communications of the ACM*, 39, pp. 55-65.

Large, A., Beheshti, J. and Rahman, T. 2002. Design criteria for children's Web portals: The users speak out. Journal of the American Society for Information Science and Technology. 53, 79-94.

Lazonder A.W. (2000) Exploring novice users' training needs in searching information on the WWW. Journal of Computer Assisted Learning 16, 326–335.

Lazonder A.W., Biemans H.J.A. & Wopereis I.G. (2000) Differences between novice and experienced users in searching information on the world wide web. Journal of the American Society for Information Science and Technology 51, 576–581.

Lee, M.W., Chen, S. Y., Chrysostomou, K. A and Liu, X. (2009) Mining Students' Behavior in Web-Based Learning Programs. Expert Systems with Applications. 36(2), 3459-3464

Lee, M. W., Chen, S. Y. and Liu, X. (2007) Mining Learners' Behavior in Accessing Web-based Interface. Lecture Note in Computer Science (LNCS). Vol. 4469, p. 336-346

Lee, C., Cheng, Y. W., Rai, S., & Depickere, A. (2005). What affect student cognitive style in the development of hypermedia learning system? Computers and Education, 44(1), 1–19.

T.S. Lee, C.C. Chiu, Y.C. Chou and C.J. Lu, Mining the customer credit using classification and regression tree and multivariate adaptive regression splines, Computational Statistics and Data Analysis 50 (2006), pp. 1113–1130.

Leong, S. and Al-Hawamdeh, S. (1999). Gender and learning attitudes in using Web-based science lessons. Information Research, 5(1).

Liaw S.S. (2002) An Internet survey for perceptions of computers and the World Wide Web: relationship, prediction, and difference. Computers in Human Behavior 18, 17–35.

Liaw, S. and Huang, H. 2006. Information retrieval from the World Wide Web: a user-focused approach based on individual experience with search engines. Computers in Human Behavior. 22(3), 501-517.

Liu, X. and Kellam, P. (2003) Mining Gene Expression Data. In: Bioinformatics: Genes, Proteins and Computers. C Orengo, D Jones and J Thornton (eds), pp. 229-244.

Liu, H., & Kešelj, V. (2007). Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users' future requests. Data and Knowledge Engineering, 61(2), 304–330

Liu, M., & Reed, W.M. (1995). The effect of hypermedia assisted instruction on second-language learning through a semantic-network-based approach. Journal of Educational Computing Research. 12(2), 159–175.

Lopez, N., Kreuseler, M. and Schumann, H. (2002). A Scalable Framework for Information Visualization, *IEEE Trans. Visualization and Computer Graphics*, 8(1), pp.39-51.

Lu, J., Yu, C.S, and Liu, C. (2003). Learning style, learning patterns, and learning performance in a WebCT-based MIS course, *Information and Management*, 40, pp. 497–507.

Lufkin, M.E. and Wiberg, M.M. (2007). Gender Equality in Career and Technical Education. In *Handbook for Achieving Gender Equity Through Education*, Ed. Klein, S.S., pp. 421-442.

Marchionini, G. (1998). Hypermedia and learning: Freedom and chaos. *Educational Technology*, 28, pp. 8–12.

Meyers-Levy, J. (1988). The Influence of Sex Roles on Judgment. *Journal of Consumer Research*, 14, 3, pp.522-530.

Minetou, C., Chen, S.Y. and Liu, X. (2008). Investigation of the Use of Navigation Tools in Web-based Learning: A Data Mining Approach. *International Journal of Human-Computer Interaction*, 24, pp. 48-67.

Mitchell, T., Chen, S.Y. and Macredie, R.D. (2005). Adapting hypermedia to cognitive styles: is it necessary? In *Proceedings of the Workshop on Individual Differences in Adaptive Hypermedia*, The Netherlands, August 23 – 26.

Modianos, D. and Hartman, J. (1990). Report on the Bradley University residential computing project, *Collegiate Microcomputer*, 8, pp. 129–37.

Morohan-Martin, J., Olinsky, A. and Schumacher, P. (1992). Gender differences in computer experience, skills and attitudes among incoming college students, *Collegiate Microcomputer*, X, 1, pp. 1-8.

Morahan-Martin, J. (1998a). The gender gap in Internet use: why men use the Internet more than women: a literature review, *Cyber Psychology and Behavior*, 1, pp. 3–10.

Morahan-Martin, J. (1998b). Males, females, and the Internet. In J. Gackenbach (Ed.), *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 169-197). San Diego, CA: Academic Press.

Moss, G., Gunn, R. and Heller, J (2006). Some men like it black, some women like it pink: consumer implications of differences in male and female website design, *Journal of Consumer behaviour*, 5, 4, pp. 328-341.

Moss, L. and Atre, S. (2003). *Business Intelligence Roadmap: The Complete Lifecycle for Decision Support Applications*. Addison-Wesley.

Nielsen, J. (2001) *Ten Usability Heuristics*. Available: http://www.useit.com/papers/heuristic.

Nielsen, J. and Norman, D. (2000). Web-site usability: Usability on the web isn't a luxury. *Information Week*. http://www.informationweek.com/773/web2.htm.

Oser, K. (2003). Marketing well to women pays off, *Direct*, December.

Palmquist, R.A., Kim, K.-S. (2000). Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science*, 51, 6, pp. 558–566.

Pask, G. (1979). *Final report of SSRC research programme HR 2708.* System Research Ltd, Richmond, Surrey

Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology*, 46, pp. 128-148.

Pask, G. (1988). Learning strategies, teaching strategies and conceptual or learning style. In R. Schmeck (Ed.), *Perspectives on individual differences, learning strategies and learning styles* (pp. 83–100). New York: Plenum.

Passig, D. and Levin, H. (1999). Gender interest differences with multimedia learning interfaces. *Computers in Human Behavior*, 15, 2, pp. 173-183.

Peterson, E. R., Deary, I. J. and Austin, E. J. (2007) Celebrating a common finding: Riding's CSA test is unreliable. *Personality and Individual Differences*, 43, pp. 2309-2312.

Pituch, K. A., & Lee, Y. K. (2006). The influence of system characteristics on E-learning use. *Computers and Education*, 47, pp. 222–244.

Potelle, H., & Rouet, J. F. (2003). Effects of content representation and readers prior knowledge on the comprehension of hypertext. *International Journal of Human–Computer Studies*, 58, 3, pp. 327–345.

Potosky, D. (2007). The Internet knowledge (iKnow) measure. *Computers in Human Behavior*, 23, 6, pp. 2760-2777.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco, CA: Morgan Kaufmann.

Raman, B. and Ioerger, T.R. (2002). Instance based filter for feature selection. *Journal of Machine Learning Research*, pp. 1–23

Reed, W.M., Oughton, J.M, Ayersman, D.J., Giessler, S.F. and Ervin, J.R. (2000). Computer experience, learning style and hypermedia navigation. *Computers in Human Behaviour*, 16, pp. 619-628.

Rezaei, A. R., Katz, L. (2004). An Evaluation of the Reliability and Validity of the Cognitive Styles Analysis Inventory and Recommendations for Improvement. *Journal of Personality and Individual Differences*. 36, 1317-1327.

Rezaei, A. R. Katz, L., (2003). An integrative approach to Collaborative Electronic Learning. *Journal of Computers in Mathematics and Science Teaching*, 22 (1), 67-83.

Richardson, A. (1977). Verbalizer-Visualizer: A cognitive style dimension. *Journal of Mental Imagery*, 1, pp. 109–126.

Riding, R. J. (1998). The nature of cognitive style. *Educational Psychology*, 17, 1-2, 29-49.

Riding, R. J. (1991). *Cognitive styles analysis*. Birmingham, UK: Learning and Training Technology.

Riding, R.J. and Al-Sanabani, S. (1998). The effect of cognitive style, age, gender and structure on the recall of prose passages, *International Journal of Educational Research*, 29, 3, pp. 173-185.

Riding, R.J., & Cheema, I. (1991). Cognitive styles—An overview and integration. *Educational Psychology*, 11, 193–215.

Riding, R.J. and Grimley, M. (1999). Cognitive Style, gender and learning from multimedia materials in 11 year old children. *British Journal of Educational Technology*, 30, 1, pp. 43-56.

Riding, R.J. and Rayner, S. G. (1998). *Cognitive styles and learning strategies*. David Fulton, London.

Riding R. J., and Sadler-Smith, E. (1992). Type of instructional material, cognitive style and learning performance. *Educational Studies*, 18, pp. 323–340.

Roussinov, D. and Zhao, J.L. (2003). Automatic discovery of similarity relationships through Web mining. *Decision Support Systems*, 35, 1, pp. 149-166.

Roy M., Chi M. T. C. (2003). Gender differences in patterns of searching the web. *Journal of Educational Computing Research*, 29, pp. 335-348.

Roy, M., Taylor, R., and Chi, M.T.H. (2004). Searching for information on-line and off-line: gender differences among middle school students. *Journal of Educational Computing Research*, 29, pp. 229–252.

Schumacher, P. and Morahan-Martin, J. (2001). Gender, Internet and computer attitudes and experiences. *Computers in Human Behavior*, 17, pp. 95–110.

Shashaani, L. (1994). Gender differences in computer experience and its influence on computer attitudes. *Journal of Educational Computing Research*, 11, 4, pp. 347-367.

Shih, C. and Gamon, J.A. (2002). Relationships among learning strategies, patterns, styles and achievement in Web-based courses. *Journal of Agricultural Education*, 43, 4.

Shih, P.C., Munoz, D. and Sanchez, F. (2006). The effect of previous experience with information and communication technologies on performance in a Web-based learning program. *Computers in Human Behaviour*, 22, 6, pp. 962-970.

Smith, B., Caputi, P., Crittenden, N., Jayasuriya, R. and Rawstorne, P. (1999). A review of the construct of computer experience. *Computers in Human Behavior*, 15, pp. 227–242.

Spink, A., Jansen, B.J., Blakely, C. and Koshman, S. 2006. A study of results overlap and uniqueness among major Web search engines. *Information Processing & Management*, 42, 5, pp. 1379-1391.

Spink, A., Wilson, T.D., Ford, N.A., Foster, A., and Ellis, D. (2002). Information seeking and mediated searching. Part I. Background and research design. *Journal of the American Society for Information Science and Technology*, 53, 9, pp. 695-703.

Sternberg, R.J., Lubart, T.I. (1996). Investing in creativity. *American Psychologist*, 51, 7, pp. 677-688.

Szpunar-Huk, E., (2006). Classifier Building by Reduction of an Ensemble of Decision Trees to a Set of Rules. *International Conference on Computational Intelligence for Modelling Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, Sydney, pp 144-148.

Treu, S. (1994). *User interface design*. Plenum Press, New York.

Urtubia, A., Perez-Correa, J.R., Soto, A. & Pszczolkowski, P. (2007). Using data mining techniques to predict industrial wine problem fermentations. *Food Control*, 18, 1, pp.1512-1517.

Weiser, E. B. (2000). Gender differences in Internet use patterns and Internet application preferences: a two sample comparison. *CyberPsychology and Behavior*, 3, pp. 167–77.

Weller H G, Repman J and Rooze G E (1994). The relationship of learning, behavior, and cognitive styles in hypermedia-based instruction: implications for design of HBI. *Computers in the Schools*, 10, 3–4, pp. 401–420.

Witkin, H.A. (1976). Cognitive style in academic performance and in teacher–student relations. In S. Messick & Associates (Eds.), *Individuality in learning* (pp. 38–72). San Francisco: Jossey-Bass.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field-dependent and field independent cognitive styles and their educational implications. *Review of Educational Research*, 47, 1, pp. 1–64.7

Yang, J. and Olafsson, S., (2006). Optimization-based feature selection with adaptive instance sampling. *Computers & Operations Research*, 33, 11, pp. 3088-3106.

Zhang, X., Gong, W. and Kawamura, Y. (2004). Customer Behavior Pattern Discovering with Web Mining, Advanced Web Technologies and Applications, In *Lecture Notes in Computer Science*, 3007/2004, pp. 844-853.

Zhu, Greiner, R., and Haubl G. (2003). Learning a model of a web user's interests. *Lecture Notes in Artificial Intelligence*, 2702/2003, pp. 11–21.

# Appendices

# Appendix A

### Dataset 1: Web Search Tools Exit Questionnaire

| Visibility of System Status |
|---|
| 1. Every page is clearly marked with a label that indicates subject contents. |
| 2. You can easily understand what options are available to you. |
| 3. There is some form of instruction to teach you how to do each activity. |
| 4. There is visual feedback when objects are selected. |
| 5. It is easy to understand which icon has been selected. |
| 6. There is always a way to get back to the home page from any other pages. |
| 7. You can easily identify where you are in the Web Site |
| 8. Navigation scheme is consistent throughout the whole program. |
| 9. The names of links and buttons are understandable. |
| 10. High contrast colour scheme is applied to present the text and background. |
| **Match Between System and the Real World** |
| 11. Menu items are arranged alphabetically. |
| 12. The colour scheme matches with your preferences. |
| 13. The length of the page matches with your expectation. |
| 14. The program avoids using computer jargon/technical terms. |
| 15. Relevant subject content appears on the same page. |
| 16. The results are presented by the levels of the relevance. |
| 17. Each topic is explained in detail. |
| 18. The subject content is divided into categories that are meaningful to you. |
| 19. From the menu list, you can see the relationships among the sub-topics presented in the menu. |
| 20. There are sufficient topics in the menu list. |
| 21. Each window has a title. |
| 22. Navigation buttons are labelled clearly and distinctively. |
| **User Control and Freedom** |
| 23. You can control your browsing paths by yourself. |
| 24. You can move back and forward between different search options. |
| 25. It is easy to pick up the relevant information directly. |
| 26. You can customise the layout, for example change the font size or link colour. |
| 27. There are sufficient links for you to link the relevant concepts. |
| 28. For menus with multiple levels, there is a mechanism that allows you to go back to the previous level. |

| Consistency and Standards |
|---|
| 29. Error messages appear in the same location throughout the system. |
| 30. Online instructions appear in a consistent location across screens. |
| 31. Icons are clearly labelled. |
| 32. There are not too many types of icons. |
| 33. Heavy use of all uppercase letters on a webpage has been avoided. |
| 34. Icons and navigation buttons are consistent from one screen to another. |
| 35. Consistent colour scheme is used in the system. |
| 36. Messages are appropriate, inoffensive, non-hostile or violent. |
| 37. When the program presents numbers, integers are right-justified and real numbers are decimal-aligned. |
| 38. Vertical and horizontal scrolling is possible in each window |
| 39. No more than four colours are utilised in each screen. |
| 40. Messages are grammatically correct.. |
| 41. Bold fonts or larger fonts are only used for emphasising important information. |
| **Help Users Recognize, Diagnose, and Recover From Errors** |
| 42. Error messages are stated constructively, without overt or implied criticism to you. |
| 43. When you are in trouble, you can receive feedback immediately. |
| 44. Messages are brief and unambiguous. |
| 45. Multiple levels of error-message details are available. |
| 46. All error messages in the system use consistent grammatical style, form, terminology, and abbreviations. |
| 47. The program highlights the error that you have made. |
| 48. Error messages let you know the cause of the problem. |
| 49. Error messages suggest what action you need to take to correct the error. |
| 50. Error messages avoid the use of exclamation points. |
| 51. There are FAQs to help you correct the errors |
| 52. Sound is used to indicate an error. |
| 53. The error messages are humorous. |
| 54. You can switch off an error message. |
| **Error Prevention** |
| 55. You will not get stuck when you make a mistake. |
| 56. The buttons that can cause serious consequences are located far away from low-consequence and high-use keys. |
| 57. When you navigate between multiple windows, the data are not lost. |

| |
|---|
| 58. Drastic and destructive consequences of your action always require your confirmation. |
| 59. Data inputs are case-blind whenever possible. |
| 60. The options that are used less frequently are located in the less-convenient positions. |
| 61. When you navigate between multiple windows, the titles are easy to identify. |
| **Recognition Rather Than Recall** |
| 62. Bright text colours are used to emphasise the most important subject and darker, duller, and de-saturated colours are used to de-emphasize the least important subject. |
| 63. White space is used to create symmetry and attract your attention to the appropriate direction. |
| 64. Sections have been separated by spaces, lines, colour, letters, bold titles, rules lines, or shaded areas. |
| 65. A map that shows the structure of the subject content is available. |
| 66. The same background colour is used to present the content within a section. |
| 67. Different colours are used to discriminate links that you have visited from links that you have not visited. |
| 68. Prompts, cues and messages are placed where the eyes are likely to be looking on the screen. |
| 69. There is a paragraph to introduce the content of each key topic. |
| 70. Different colours are used to differentiate different subject topics. |
| **Flexibility and Efficiency of use** |
| 71. Multiple levels of detail are applied to illustrate the subject content. |
| 72. The program provides short cuts to access detailed information directly. |
| 73. There is a detailed alphabetical index to help you locate specific information. |
| 74. The system allows you to decide what topics should be looked at first. |
| 75. The search engine provides you with multiple search options. |
| **Aesthetic and Minimalist Design** |
| 76. Icons of different groups are visually distinctive. |
| 77. Only information essential to the task is displayed on the screen. |
| 78. The title of each page is short, simple, clear, and distinctive. |
| 79. Menu items are brief, yet long enough to describe the subject content. |
| 80. There are some nice pictures to decorate the layout. |
| 81. Relevant icons and buttons are grouped together. |
| **Help and Documentation** |
| 82. Online instructions are visually distinctive |
| 83. It is easy for you to find the HELP button. |
| 84. The layout of the online help is consistent with the main text. |

| |
|---|
| 85. The help provides you with step-by-step procedures. |
| 86. The help can provide you with enough information to illustrate common errors. |
| 87. You can easily switch between the help system and your work. |
| 88. Help is presented as pop-up window to let you do your task and consult the help system at the same time. |
| 89. There is an alphabetical index to help you to locate information within the help system. |
| 90. It is easy to access and return from the help system. |

# Appendix B

## Dataset 2: Web-Based Instruction Exit Questionnaire

| # | Question |
|---|---|
| 1 | It is difficult to learn the basics of HTML using this tutorial without the help of a person. |
| 2 | Examples given in this tutorial are not practical. |
| 3 | I felt the structure of this tutorial is not clear. |
| 4 | I sometimes got lost because the buttons made me feel confused. |
| 5 | I spent a lot of time getting to know how to use this tutorial. |
| 6 | I would have found it more helpful to be given a suggested route through this tutorial. |
| 7 | I would like to have more examples. |
| 8 | I would prefer to learn from human tutor than from this tutorial. |
| 9 | It is hard to use back/forward buttons. |
| 10 | The information provided by the map is too superficial. |
| 11 | The links provided in this tutorial help me discover relationships between different topics |
| 12 | The map in this tutorial gives a meaningful framework of HTML. |
| 13 | I was confused which options I wanted, because it provided too many choices. |
| 14 | After using this system I can easily use my knowledge to design home pages. |
| 15 | I found it hard to select relevant information using the map. |
| 16 | I like the fact that it allowed me to learn topics in any order. |
| 17 | I like the fact that this tutorial allowed me to work at my own pace and direction. |
| 18 | It is easy to find a route for a specific task with the index. |
| 19 | This tutorial can be used sufficiently well without any instructions. |
| 20 | I felt difficult to browse pages containing texts and links in the same pages. |