INFORMATION QUALITY ASSESSMENT IN E-LEARNING SYSTEMS

Mona Awad ALKHATTABI

Thesis submitted for the degree of

Doctor of Philosophy in Computer Science

Informatics Research Institute

School of Computing, Informatics and Media

University of Bradford

2010

Abstract

Keywords: E-learning, Information quality, Information system, Information extraction, Quality framework, Quality metrics.

E-learning systems provide a promising solution as an information exchanging channel. Improved technology could mean faster and easier access to information but does not necessarily ensure the quality of this information. Therefore it is essential to develop valid and reliable methods of quality measurement and carry out careful information quality evaluations.

Information quality frameworks are developed to measure the quality of information systems, generally from the designers' viewpoint. The recent proliferation of e-services, and e-learning particularly, raises the need for a new quality framework in the context of e-learning systems. The main contribution of this thesis is to propose a new information quality framework, with 14 information quality attributes grouped in three quality dimensions: intrinsic, contextual representation and accessibility. We report results based on original questionnaire data and factor analysis. Moreover, we validate the proposed framework using an empirical approach. We report our validation results on the basis of data collected from an original questionnaire and structural equation modeling (SEM) analysis, confirmatory factor analysis (CFA) in particular.

However, it is difficult to measure information quality in an e-learning context because the concept of information quality is complex and it is expected that the measurements will be multidimensional in nature. Reliable measures need to be obtained in a systematic way, whilst considering the purpose of the measurement. Therefore, we start by adopting a Goal Question Metrics (GQM) approach to develop a set of quality metrics for the identified quality attributes within the proposed framework. We then define an assessment model and measurement scheme, based on a multi element analysis technique. The obtained results can be considered to be promising and positive, and revealed that the framework and assessment scheme could give good predictions for information quality within e-learning context.

This research generates novel contributions as it proposes a solution to the problems raised from the absence of consensus regarding evaluation standards and methods for measuring information quality within an e-learning context. Also, it anticipates the feasibility of taking advantage of web mining techniques to automate the retrieval process of the information required for quality measurement. This assessment model is useful to e-learning systems designers, providers and users as it gives a comprehensive indication of the quality of information in such systems, and also facilitates the evaluation, allows comparisons and analysis of information quality.

Dedicated to my parents, my husband and my lovely children

*for filling my life with joy and happiness***.**

Peer Reviewed Publications and Contributions

Journal Publications:

- M. Alkhattabi, D. Neagu, and A. Cullen, Assessing information quality of e-learning systems, *"Computers in Human Behavior"* Elsevier, **Under Review**
- M. Alkhattabi, D. Neagu, and A. Cullen, "Information quality framework for e-learning systems," *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 2010 **Accepted**

Chapter in Book:

- M. Alkhattabi, D. Neagu, and A. Cullen, "User perceptions of information quality in e-learning systems: A gender and cultural perspective," chapter in *Globalization, technology diffusion and gender disparity: Social impacts of ICTs* R. Pande, T.V.D. Weide, and N. Flipsen, IGI Global 2011, **Accepted**

Conference Contributions:

- M. Alkhattabi, A. Cullen, and D. Neagu, "Information quality metrics in an e-learning context: Applying GQM approach," *Proceedings of 7th International Conference on the Quality of Information and Communications Technology*, Oporto, Portugal 29 September to 2 October 2010. **Accepted**
- M. Alkhattabi, D. Neagu, and A. Cullen, "Information quality in information systems: A critique of relevant literature," *Proceedings* of *the 4th Saudi International Conference* Manchester, 30-31 July 2010.
- M. Alkhattabi, D. Neagu, and A. Cullen, "Quality in e-learning systems: An overview of the definition and concept," *Proceedings* of the *4th Saudi International Conference* Manchester, 30-31 July 2010.
- M. Alkhattabi, D. Neagu, and A. Cullen, "Validating an information quality framework for e-learning systems: An empirical approach," *Proceedings of the 15th International Conference on Software Process Improvement Research, Education and Training, INSPIRE (e-Learning and Social Responsibility)*, 47-54, M. Ross, Ed. BCS, London, 31st of March 2010
- M. Alkhattabi, D. Neagu, and A. Cullen, "Semantic web mining for quality information used in e-learning," *Proceedings of the 9th Informatics Workshop*, 56-59, University of Bradford, UK, June, 2008

Presentations:

- The Institution of Engineering and Technology (IET) prestige invited talk: Information quality in e-learning, with Dr Daniel Neagu and Dr. Andrea Cullen, (2009)

Table of Contents

Tables

Figures

Acknowledgments

Firstly, I would like sincerely to thank ALLAH, the most merciful and the most gracious, who makes all things possible.

My parents deserve a special mention for their inseparable support and prayers. To my father, who taught me how to dream, thank you for being so supportive. Many thanks go, in particular, to my mother who sincerely raised me.

I give loving thanks to my husband (Fahad) who shared with me my dreams and whose dedication and persistent confidence in me have taken the load off my shoulders. To say “thank you” just seems completely inadequate. A special thank you goes to my three little heroes (Mohammed, Dema and Sarah), for your love, patience and confidence in me all the time, sharing both ups and downs in my study. How could I ever have done this without you?!

This thesis could not have been completed without Dr. Daniel Neagu and Dr. Andrea Cullen who not only served as my research supervisors, but also encouraged and challenged me throughout my academic journey. They never accepted less than my best efforts. Thank you.

I would like to record my appreciation to staff of both the Ministry of Higher Education and the Saudi Arabian Cultural Bureau for the supervision, advice and guidance they provided from the very early stage of this work.

Finally, I would like to note my gratitude for all those who helped in collecting data and also the people who agreed to take part in the research.

List of Abbreviations

AC	Accessibility Information Quality Factor
ALN	Asynchronous Learning Network
AMOS	Analysis of Moment Structures
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CIS	Cooperative Information System
CR	Contextual Representation Information Quality Factor
D&M	DeLone and McLean
DWQ	Data Warehouse Quality
E-learning	Electronic Learning
GFI	Goodness of Fit Index
GQM	Goal Question Metrics
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
IE	Information Extraction
InT	Intrinsic Information Quality Factor
IQ	Information Quality
IS	Information System
JAXP	Java API for XML Processing
J2SE	Java 2 Standard Edition

KMO	Kaiser-Meyer-Oklin
LSP	Learning Service Provider
NL	Natural Language
RSMEA	Root Mean Square Error of Approximation
SEM	Structure Equation Modelling
SPSS	Statistical Package for the Social Science
TLI	Tucker Lewis Index
UML	Unified Modelling Language
VLE	Virtual Learning Environment
WWW	World Wide Web
XML	eXtensible Markup Language

Chapter 1: Introduction

1.1 Introduction

The recent widespread global use of the Internet and the World Wide Web (WWW) has resulted in remarkable opportunities for the development of different electronic services. As a result, the area of e-services is becoming extremely attractive for the researcher to explore the possibilities for extensive technological and economical impact, opening new domains like e-learning, e-government, e-commerce, e-advertising, social networking etc.

E-learning, as one of the main areas of e-services, has undergone intensive development as an inevitable result of the recent proliferation of Internet technology. Traditional means of learning restrict the learner to certain learning methods, at a specific time and place; whereas, e-learning services create wider horizons for organisations and individuals who are involved in teaching and learning. These environments facilitate the delivery of substantial parts of education through the use of tools and materials that are accessible directly in the learner's home or office, and at any time. In addition, the advancements in technology, which are used to enhance the interactivity and media content of the web and the increasing quality of delivery platforms, create an ideal environment for the expansion of e-learning systems.

However, problems could ensue as a result of the remarkable increased reliance on e-learning systems at different levels across education. The growing number of available e-learning systems and the commercialisation of these systems highlight the necessity of quality evaluations of online published learning materials. Although quality evaluation of learning materials in e-learning systems have become increasingly important, the actual evaluation standards and methods for Information Quality (IQ) in such systems have not yet reached a consensus.

This study focuses on the evaluation of IQ in the context of e-learning systems, and in particular, on identifying the key dimensions for IQ from the users' perspective, in order to build a framework for IQ in e-learning systems. The study will identify a set of quality metrics to quantify the quality of the information. This will be based on the proposed framework and will ultimately facilitate the evaluation, allow comparison, and analysis of IQ. Furthermore, it will investigate the possibility of integrating a web mining approach, information extraction technique, in order to automate the evaluation process.

The purpose of this chapter is to present an overview of the research problem, the rationale, and the organisational context for the study; in addition, a brief outline of the thesis will be provided.

1.2 Motivation

E-learning services and technologies provide learners with distributive, collaborative, and interactive features which help to overcome the restrictions of space and time, this permits information to be delivered and received, and allows the learners to create their own learning paths and procedures. In fact, e-learning has progressed from simply a delivery tool to a fundamental learning mechanism for the whole learning process. Nevertheless, the lack of direct contact with learners, which is considered to be a significant obstacle in determining the effectiveness of the educational process, raises many questions about the overall quality of the educational outcomes. Literature in this field examines the quality of the published materials in two ways: firstly, focusing on the content of these materials, and secondly examining the way in which such content will be delivered to the students. In fact, the measurement of the quality of content delivered by e-learning systems is found to be the most important and most influential factor in evaluating the overall quality of the learning[1].

In itself, quality is a very abstract notion; it is difficult to provide a simple definition of IQ. In fact, taking into account that quality, on the web, is a complex concept and its measurement is expected to be multidimensional in nature [2], identifying the criteria, by which the quality is determined, is considered to be the prime issue in evaluating the quality of any online distributed system [3]. These evaluation criteria are a result of the multidimensional and interdependent nature of quality in the distributed systems, and are therefore dependent on the objectives and context of the system. Moreover, the specified context and the perspectives of the users need to be considered when defining quality in an e-learning context.

Despite the quality evaluation of learning materials being a critical issue, there is still no consensus regarding evaluation standards and methods for measuring IQ in e-learning systems. Furthermore, the criteria and methods utilised to evaluate e-learning materials and systems should have specific characteristics which differ from the methods used to evaluate typical learning materials. To solve this problem, comprehensive and specific quality criteria are needed; as such, measurement metrics with clear quality benchmarks must be developed and a suitable and reliable method, to apply these criteria to given e-learning materials, must be selected [4].

[1] As will be demonstrated in Chapter 2

The majority of methodologies which are used to evaluate the quality of the distributed learning materials are qualitative and are generally user oriented because they depend on linguistic recommendations from the students. Moreover, the measurement process is dependent on the learner's perspective, since the selected criteria must be easily comprehensible by the learner. Consequently, the measurement schemes are user centred because they produce linguistic recommendations of the learning material, these are based on the learner's linguistic estimation judgments. The main problem with these traditional measurement methods is that they rely on human judgment, this can be uncertain and inaccurate, and also entails large amounts of effort and time. It therefore seems logical to take advantage of web mining techniques to automate the retrieval process of the information needed in the quality measurement.

1.3 Objective

This study will focus on proposing measures for the quality of the content provided by e-learning systems by identifying the main quality standards. The major issue in measuring the quality for any online distributed systems is identifying the criteria of the quality, as the quality in distributed online systems, such as e-learning systems, is considered to be a multidimensional and interdependent subject that is dependent on the objectives and context of the system. Moreover, because web quality is a complex concept, its measurement is neither simple task nor straightforward in fact it expected to be multidimensional in nature [5].

This research will present quality evaluation metrics to measure the content of the learning materials distributed via e-learning systems. Moreover, because human judgment is fallible, this research will aim to not only build suitable measurement metrics to characterise in a deterministic way the quality of the content provided by distributed learning materials, but it will also focus on the feasibility of integrating web mining techniques as a means of gathering the necessary information to conduct the evaluation measurement.

In general, once the main dimensions of the IQ are defined – which set the quality standards for the content provided by e-learning system, the focus will be on identifying a set of metrics to quantify the quality of the information – this will facilitate the evaluation, comparison, and analysis of IQ. Moreover, it will focus on the development of the assessment scheme. This work will identify the appropriate data collection techniques, tools and procedures to allow the assessment process to become automated. By analysing the IQ from a given e-learning system, the collected data, the quality scores, and the feedback can provide suggestions and recommendations for future improvements.

1.4 Research Questions

As previously stated, this study will focus on measuring the quality of the content provided via web based learning systems, it will identify the main criteria used to determine quality and will build a quality framework, for IQ, in the context of e-learning systems. Afterwards, a set of quality metrics will be identified to quantify the quality of the information based on the proposed framework. Furthermore, it will examine the possibility of automating the process of collecting the required data for the defined quality metrics. Accordingly, the focus of this thesis is on the following four research questions:

– ***Research Question (1)***

How can the key dimensions for IQ be identified, from the users' perspective, in order to build a quality framework to measure the quality of the content provided by e-learning systems?

– ***Research Question (2)***

How can a specified set of metrics be determined, to quantify the quality of information in e-learning systems, in such way they will enable the evaluation, comparison, and analysis of IQ in such systems?

– ***Research Question (3)***

What is the most appropriate and applicable assessment scheme, used to compute the identified metrics, and which will ultimately reach an overall IQ assessment for the published materials?

– ***Research Question (4)***

How can the web mining technologies be positively utilised, in order to automate the data collection and evaluation processes?

1.5 Research Approach and Methodology

In order to achieve the research objectives, the research was structured in the following three phases:

1- Development and validation of an IQ framework for e-learning systems;
2- Derive quality metrics and define a suitable assessment scheme;
3- Testing of the framework, quality metrics and the measurement scheme through a case study.

The first phase involved developing and validating a framework which would represent possible factors and attributes that might impact upon IQ in e-learning systems. The earlier proposed frameworks, from the relevant literature, were used together with a user questionnaire in order to build the framework. The questionnaire was designed according to the methods introduced by Churchill [6], and involved a cross-sectional survey and a sample of 315 participants. The aim was to determine the users' view of the relative importance of the quality dimensions for information published in e-learning systems. The collected data were used along with factor analysis and linear regression to build the quality framework.

For the validation process, a more focused and purpose-driven survey was conducted. One hundred responses were collected from four academic institutions in Saudi Arabia; all of the selected institutions had implemented enterprise e-learning systems. The proposed framework was validated using a confirmatory factor analysis (CFA) approach [7] as a mean of structural equation modelling (SEM) goodness-of-fit test [8].

After that, a goal question metric (GQM) approach [9] was used in the metric identification, to quantify the quality of the information, in order to facilitate the evaluation, comparison, and analysis of the IQ. Then, a multi element analysis technique was used to define the assessment process in the next phase [10].

For the third phase, an automated approach was applied, using a web mining technique, as a feasibility test for the proposed approach. The achieved experimental results were used to compute quality scores and feedback about the quality of the information within a given e-learning system. Figure 1 shows the structure of the thesis and the steps followed throughout the research.

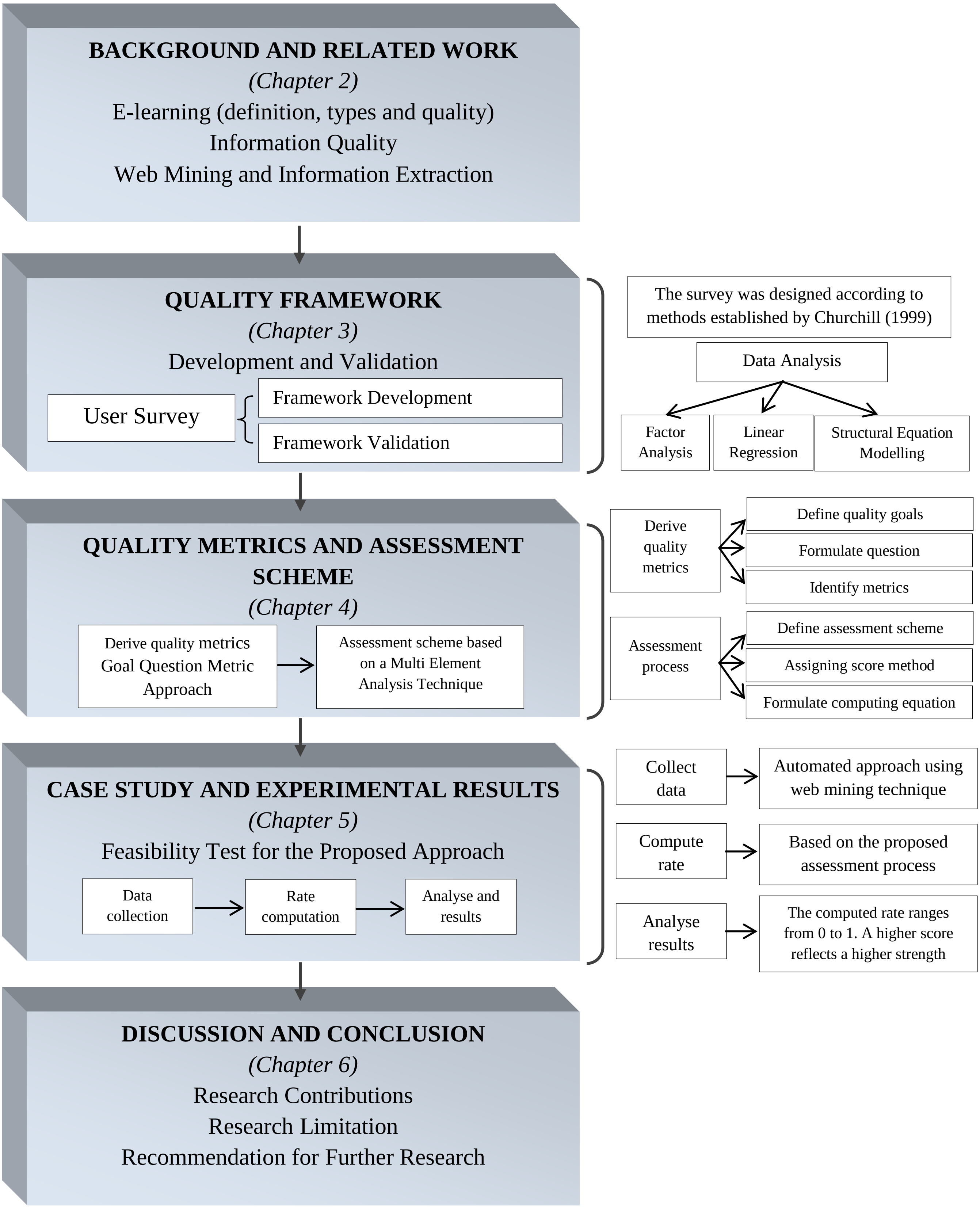


Figure 1. The structure of the thesis and steps followed in the research method

The previous figure will be used as a starting point for each chapter throughout the thesis to clarify role of the chapter within the general plan of the thesis and linking it with the previous and the remaining steps.

1.6 Outline of the Thesis

This thesis comprises of six chapters:

- ***Chapter 1:*** provides the background to the research and introduces the research questions and a brief overview of the research approach and methodology. Finally, the layout and content of the chapters are described.
- ***Chapter 2:*** reviews literature from the three main disciplines of this research: this encompasses quality concepts in e-learning, information quality in information system (IS), and web mining and information extraction.
- ***Chapter 3:*** focuses on our contributions towards identifying the key dimensions for IQ from the users' perspective in order to build a quality framework for measuring the quality of the content provided by e-learning systems. Also, it gives a brief idea of how the framework could be used as a means to examine the differences in users' perceptions of information quality in e-learning systems from a gender and cultural perspective. It then presents an empirical validation of the proposed quality framework.
- ***Chapter 4:*** presents our work for quality metrics identification procedure and uses a goal question metric (GQM) approach in order to quantify the quality of the information. It also discusses our work to define the assessment scheme, based on a multi element analysis technique.

- ***Chapter 5:*** discusses the case study and the experimental results; an automated approach was applied using a web mining technique as a feasibility test for the proposed approach. In addition, quality scores were computed for the IQ within specific e-learning system, the results from the case study were used to achieve this.
- ***Chapter 6:*** presents the major conclusions of the research as well as any research contributions. Finally, the limitations of this research are discussed, along with suggestions for future research.

1.7 Summary

This chapter lays the foundation for the research by providing background information and by introducing the research and the research aims and objectives. The research approach and methodology were then presented, finally, an outline of the thesis was provided. The next chapter, Chapter 2, will present a detailed literature review.

Chapter 2: Background and Related Work

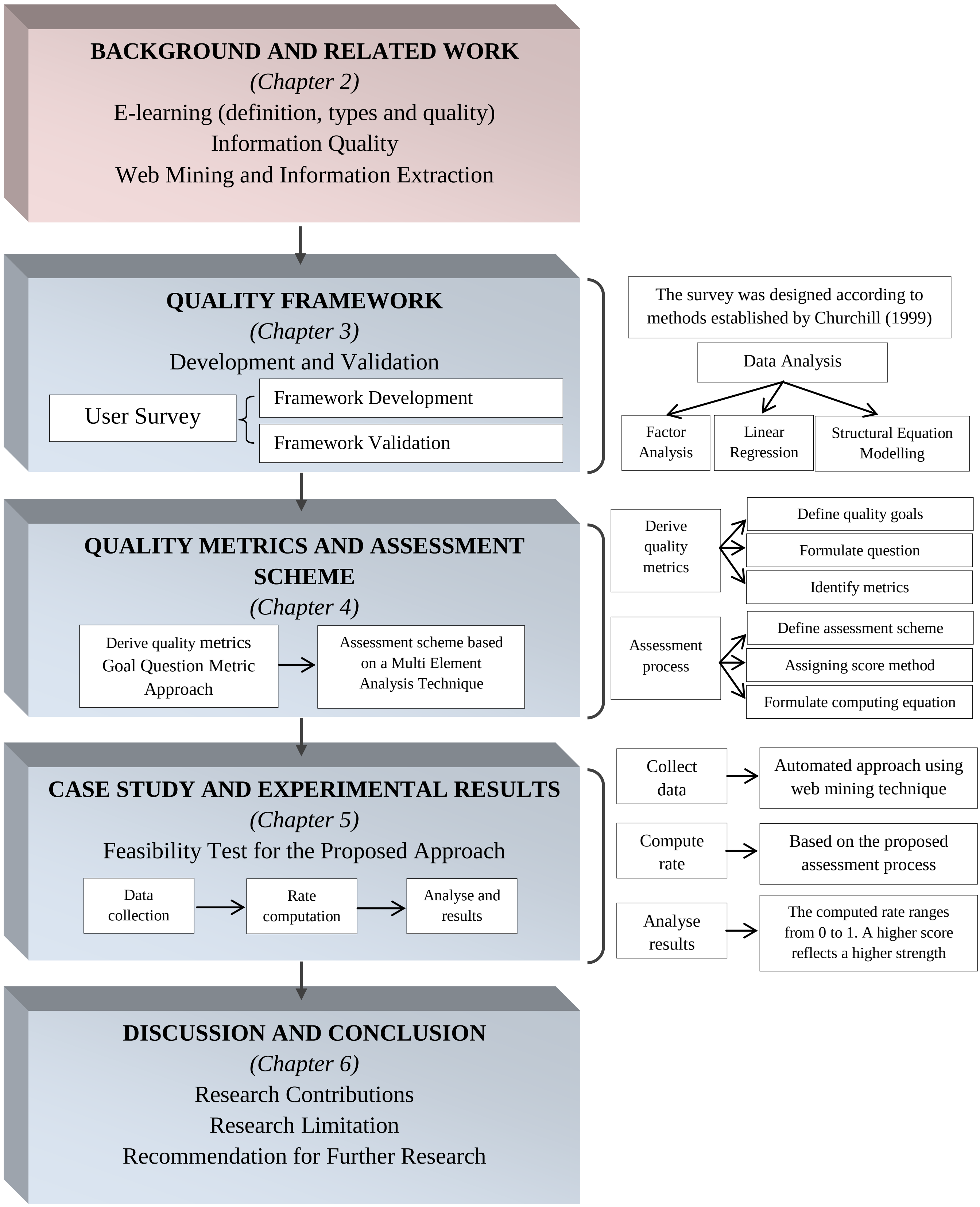


Figure 2. Thesis structure / Chapter 2

2.1 Introduction

During the last decade the amount of literature published in the field of e-learning has grown noticeably, as has the diversity in attitudes and viewpoints of people who work on this subject. The general background presented here with regard to e-learning includes the definition, details of different types and the concept of quality. Information quality within information systems, web mining and information extracting techniques are the main areas on which supporting literature is primarily focused. However, an in-depth explanation of each branch of these research fields is outside the scope of this literature review. The literature presented here is particularly focused on the subtopics of these large research areas which are directly applicable to this research.

The structure of this chapter is divided into three main parts: a general view of e-learning including definitions, types and the concept of quality; information quality within information systems; and information extraction methods. Each section includes a number of subsections which address the factors that are relevant to this research.

2.2 E-learning

In this part of the literature review, we focus on providing a discussion aiming to reach a clearer understanding of e-learning definitions, types and the concept of quality. Moreover, in this section we lay the foundation for the general concept of information quality in an e-learning context, upon which the research will be based. This section also presents a discussion about the relationships between technology, users and content within e-learning systems.

2.2.1 E-learning Definition

The term e-learning is used in the literature and in business to describe many fields, such as online learning, web-based training, distance learning, distributed learning, virtual learning, or technology-based training [11]. During recent decades, e-learning has been defined in several instances in different ways. In any publication in the field of e-learning, it is important to ensure that the author's understanding exactly matches that of the majority of the readers, therefore, the specific definition used should be stated first [12]. Moreover, to reach a clearer understanding of what e-learning is, in this part of the thesis we present various definitions as mentioned in the literature.

In general, most of the definitions of the term e-learning are used to express the exploitation of technologies which can be used to deliver learning (or learning materials) in an electronic format, most likely via the world wide web [13]. In the same context, Psaromiligkos and Retalis [14] consider that e-learning comprise of systems which utilise the world wide web as a delivery medium for static learning resources, such as instructional files, or as an interface onto interactive.

The previous definitions look at e-learning in general; in more detail, e-learning can be in the form of courses or in the form of modules and smaller learning materials – it also could take various forms. Romiszowski [12] takes these details into account and summarises the definitions encountered in the literature in a way that emphasises that e-learning can be a solitary, individual activity, or a collaborative group activity. It also suggests that both synchronous and asynchronous interactive forms can be engaged (which will be explained in more details in the next section). Naidu [15] also takes into consideration the differences in the forms of interaction when trying to formulate a general definition of e-learning:

> *"… educational processes that utilize information and communications technology to mediate asynchronous as well as synchronous learning and teaching activities." [15 p.1]*

The position adopted in this research is that e-learning entails the technology used to distribute the learning materials, the quality of these materials, and the interaction with learners. The definition of e-learning used in this research addresses these dimensions in terms of:

> *"… the use of new multimedia technologies and the Internet to improve the quality of learning by facilitating access to resources and services as well as remote exchange and collaborations" [16 p. 2]*

2.2.2 E-learning Types

As mention earlier, e-learning takes many different forms and includes numerous types of systems. In the existing literature e-learning types are defined following two main axes: the user context (individuals, groups or a community of users) and users' engagement and interactivity [17].

As mentioned earlier, Romiszowski [12] takes these details into account and summarises them in his structured definition of e-learning in Table 1.

He emphasises that e-learning can be a solitary, individual activity, or a collaborative group activity. It also suggests that both synchronous and asynchronous interactive forms can be engaged [12]. Looking more deeply at the division of the forms of interactivity used in e-learning systems, there are two main types: asynchronous and synchronous, depending on learning and teaching activities [18].

	(A) ***INDIVIDUAL SELF-STUDY*** ***Computer-based instruction/ learning/training (CBI/L/T)***	***(B)*** ***GROUP COLLABORATIVE*** ***Computer-mediated communication (CMC)***
(1) ***ONLINE STUDY*** ***Synchronous communication*** ***("REAL-TIME")***	Surfing the Internet, accessing websites to obtain information or to learn (knowledge or skill) (Following up a WebQuest)	Chat rooms with(out) video (IRC; electronic whiteboards) audio/videoconferencing (CUSeeMe; NetMeeting)
(2) ***OFFLINE STUDY*** ***Asynchronous communication*** ***("FLEXI-TIME")***	Using stand-alone courseware/ downloading materials from the Internet for later local study (LOD-learning object download)	Asynchronous communication by e-mail, discussion lists or a learning management system (WebCT; Blackboard; etc.)

Table 1. Structured definition of e-learning. Source [12 p.6]

While, synchronous e-learning environments require tutors and learners, or the online classmates, to be online at the same time, where live interactions take place between them, asynchronous learning network (ALN) could be described as a place where learners can interact with learning materials, tutors and other learners, through the WWW at different times and from different places [19]. Moreover, Doherty [20] describes ALN as a variety of e-learning systems which distribute learning materials and concepts in one direction at a time.

The focus of this research will be on a case where students log-in to and use the system independently of other students and staff members, as well as using asynchronous methods regarding learning content, quality management and delivery which fit firmly into the general definition of the asynchronous e-learning environment.

2.2.3 Concept of Quality in E-learning Context

The definition of e-learning adopted in this thesis represents three fundamental dimensions: technology, access and quality. The focus in this research will be on quality, which is considered a crucial issue for education in general, and for e-learning in particular. This section of the literature review will discuss concepts of quality in e-learning generally, and highlight the importance of content as the most critical factor for the overall quality.

Currently, there are two recognised challenges in e-learning: the demand for overall interoperability and the request for (high) quality. However, quality cannot be expressed and set by a simple definition, since in itself quality is a very abstract notion. In fact, it is much easier to notice the absence of quality than its presence [21].

Despite efforts to reach a comprehensive, universal definition of quality in e-learning, there is still a fundamental ambiguity surrounding the issue. One position is to consider quality as an evaluation of excellence, a stance which is primarily adopted by universities and education institutions. [22]. While, another trend is to consider the improvement in quality, where quality is improved by moving beyond the set conceptions applied, and generally moving in the direction of a flexible process of negotiation, which needs a very high level of quality capability from those involved [23].

Furthermore, quality can be viewed and considered from different aspects. Here, the SunTrust Equitable report [1] illustrates what they perceive to be the value chain in commercials e-learning packages in the form of a pyramid. Figure 3 shows the suggested value chain pyramid where the content is the most critical factor of e-learning. Indeed, to be able to use the Internet as a tool to improve learning, the content should not distract learners, but increase their interest for learning. Learning tools and enablers are also important in the learning procedure. In reality, providers of learning platforms and knowledge management systems are key in the successful delivery of content. These companies provide the necessary infrastructure to deliver learning content. Moreover, learning service providers (LSP) are the distribution channels for content providers. One of the challenges facing these knowledge hubs and LSP is to ensure that the learners are receiving fresh content. Companies focused on educational e-tailing then complete the value pyramid of e-learning.

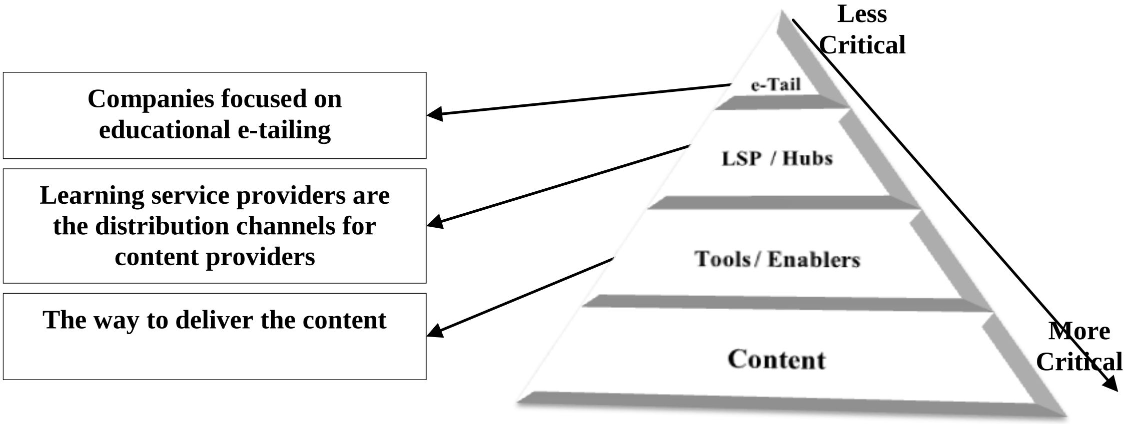


Figure 3. E-learning value pyramid. Source [1 p.11]

Looking at the pyramid it can be clearly observed that content is the most critical component of learning through the Internet [1]. In a similar manner, Henry [24] stated that e-learning is composed of three main aspects: content, technology and services, he also emphasised that content is the most significant factor.

Although this thesis will focus on the quality of content as the most important criteria and the most influential in the overall level of learning quality for any e-learning system, it is worth mentioning than there are two main types of e-learning packages: commercial and specific educational tools. Moreover, since each type could target different audience we suggested more future work to examine influence of these differences on the perspective of quality.

It is also important to specify the context and the perspectives of users when defining quality in e-learning. It is also essential to classify suitable criteria to address this quality [3].

2.2.4 E-learning Technology, Users and Content

Although most e-learning definitions focus on the technology and not on the learning, it is important to emphasise that individual learning goals, styles and required learning materials should be addressed first. Then a suitable electronic delivery method can be adopted. On their website, www.agelesslearner.com, Karl and Marcia Conner commented, in this regard, that:

> *“.....Maybe the ‘e’ should actually follow the word ‘learning’” [25]*

Henry [24] describes the content in a way that includes all delivered materials, including the materials which are usually offered in classroom based learning and that are tailored for e-learning, in addition to any other knowledge the developer might offer.

In fact, e-learning systems are considered to be user-adaptive systems, where systems are designed to react with user performance and choices. Moreover, Heift and Nicholson [26] believe that e-learning systems as adaptive systems are designed to meet the diverse requirements of students who have different levels of knowledge and backgrounds.

There is a significant base of literature and research in the area of adaptive systems, which usually base their behaviour on user models. In more detail, Kobsa [27] explained that the user model often depends on one user or a group of users sharing the same profile and it characterises user's preferences, goals, interests and knowledge.

Although e-learning systems are considered types of adaptive systems, the difference between the concept of the user and the concept of the student creates a fundamental problem in the e-learning area. In this context, Esposito, Licchelli and Semeraro [28] believe that in a general web system the user is free to surf and the system attempts to predict future user steps using the user model in order to improve the interaction between the user and the system. While in the e-learning system the modelling has to improve the educational route, adapting it to the model of the student.

Although delivering web based educational materials can be very useful as the same content is distributed to a number of students and can be accessed regardless of time and place, this delivery would not be beneficial from a pedagogical point of view if the students, their level of knowledge and their learning style was not known. In fact, Sanatally and Senteni [29] observe that the widely held principle of using the web simply as a form of distributed medium for learning materials does not add significant value to the learning process. This argument leads to the conviction of the importance of developing adaptive e-learning systems. Even if adaptive systems are focused on the interaction with users and changing the course and the content dynamically with their needs, and not on controlling the set sequence of a course, e-learning can exploit adaptive technologies to build learning environments that form user-specific sequencing.

In the case of this research, the student and domain model did not entail the complexity of those built in adaptive systems. In fact, because the focus in this research is on measuring the quality of the content of learning materials distributed via e-learning systems, and establishing the characteristics and factors with the most impact on the level of quality from the users' viewpoint, we will gather empirical evidence using online questionnaires, which can be used to directly ask students about their preferences and perspectives.

2.2.5 Summary

This part of the literature review provided a general overview of e-learning, including definitions of e-learning, a note of e-learning types and consideration of the concept of quality in e-learning. It also identified the definition adopted for e-learning in this study and considered the type upon which this research will focus. Moreover, in this section we laid the foundation for the general concept of quality in e-learning upon which the research will be based. Finally, it presented a brief discussion about the relationships between technology, users and content in an e-learning context.

The next part of this chapter will discuss the concept of IQ within information systems; this will be used later on to set standards for IQ in the context of e-learning systems.

2.3 Information Quality in Information Systems

In this part of the literature review we will start with a brief discussion of the terms "data quality" and "information quality", and will shed some light on the concept of IQ within information systems and how it could be defined. We will also provide a comprehensive review of the major historical developments of IQ frameworks.

2.3.1 Data Quality vs. Information Quality

During recent years, much work has been done to build quality frameworks for IQ dimensions. In the past, research focused on data quality (DQ), but due to the recent development of Internet technologies, information systems today are providing users with information, not only data. Therefore, research attention has shifted to focus on IQ frameworks.

While, some researchers explicitly distinguish between the terms "data" and "information" and explain information as data which has been processed in some way [30, 31], sometimes, it may be difficult to discriminate between them in practice [31].

Still, in some studies the term "information" is interchangeable with "data" [32]. Likewise, the term "data quality" is often used synonymously with "information quality" [33]. Consequently, in this study, the concept of information will be used in a broad sense, which covers the concept of data.

Before reviewing the researches that were conducted to formulate (data/information) quality frameworks within information systems, first we will discuss the meaning of IQ and how it could be defined.

2.3.2 Definition of Information Quality

Although it is important to set standards for IQ, it is a difficult and complex issue, particularly in the area of information systems, because there is no formal definition of IQ, as quality is dependent on the criteria applied to it. Furthermore, it is dependent on the targets, the environment and from which viewpoint we look at the IQ, that is, from the provider or the consumer perspective. Moreover, IQ is both a task-dependent and a subjective concept. Juran [34] summarises these aspects of quality in his quality definition as “fitness for use”. Similarly, Wang [35] described DQ (which could apply to IQ) as data that is fit-for-use.

This description has been adopted by researchers because it brings to light the fact that IQ cannot be defined and evaluated without knowing its context [36]. Defining IQ in a contextual approach seems to be logical because quality criteria, which could be used to assess IQ, can differ according to the context [37]. In fact, IQ is expressed in the literature to be a multidimensional concept with varying attributed characteristics depending on the context of the information [38]. However, taking into account the complexity of IQ concept and that its measurement is expected to be multidimensional in nature [5], the prime issue in defining the quality of any IS is identifying the criteria by which the quality is determined [2]. The criteria result from the multidimensional and interdependent nature of quality in information systems, and are dependent on the objectives and the context of the system. Thus, it is common to define IQ on the Internet by identifying the main dimensions of the quality, for that purpose IQ frameworks are widely used to identify the important quality dimensions in a specific context, these dimensions can be used as benchmark to improve the effectiveness of information systems, as described by Porter [39].

2.3.3 Information Quality Frameworks

Today, for any IS to be judged successfully it has first to satisfy additional predefined quality criteria [40]. An e-learning system is a special type of IS so it is important to examine the literature relating to the traditional IS success models and the proposed quality frameworks, in order to test the possibility of extending these success models to identify content quality criteria in an e-learning context.

Much of the work done in IS success has its origins in the well-known DeLone and McLean (D&M) IS Success Model [41]. This model provided a comprehensive taxonomy on IS success based on the analysis of more than 180 studies on IS success and it identified over 100 IS success measures during the analysis. It established that system quality, information quality, use, user satisfaction, individual and organisational impact were the most distinct elements of the IS success equation. In a later work, the authors confirmed the original taxonomy and their conclusion, namely that IS success was an interdependent and multidimensional construct [42].

Their model makes two important contributions to the understanding of IS success. First, it provides a scheme for categorising the multitude of IS success measures that have been used in the literature. Second, it suggests a model of temporal and causal interdependencies between the categories. The updated model, which was proposed in 2003, consists of six dimensions [42]:

- ***Information quality***, which concerns the system content issue. Web content should be personalised, complete, relevant, easy to understand and secure.
- ***System quality***, which measures the desired characteristics of a web based system such as usability, availability, reliability and adaptability.
- ***Service quality***, which measures the dimensions of service quality such as tangibles, reliability and assurance.
- ***Usage,*** which measures visits to a website, navigation within the site and information retrieval.

– ***User satisfaction***, which measures user's opinions of the system and should cover the entire user experience cycle.

– ***Net benefits***, which capture the balance of positive and negative impacts of the system on the users. Although this success measure is very important, it cannot be analysed and understood without system quality and IQ measurements.

The model is presented in Figure 4:

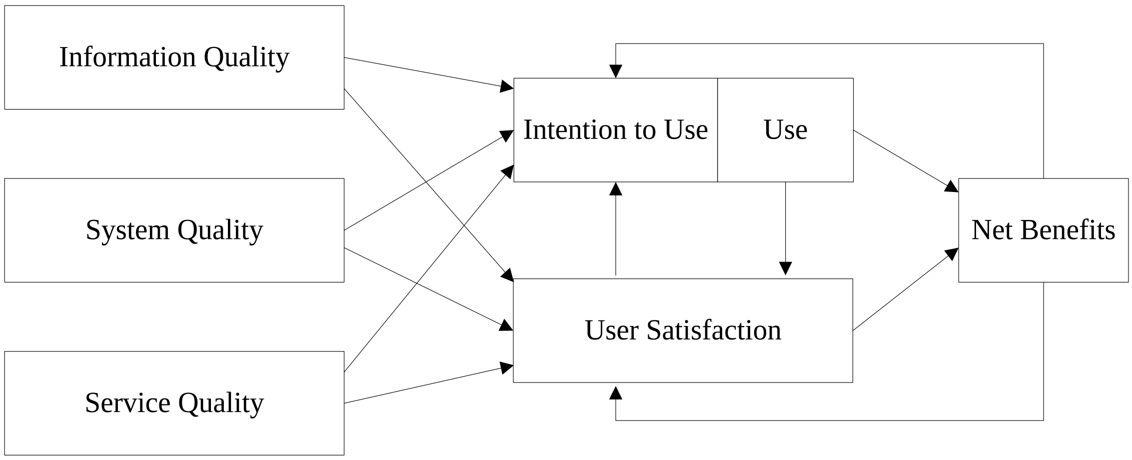


Figure 4. DeLone & McLean updated information system success model. Source [42 p.24]

In their model, DeLone and McLean defined three main dimensions for the quality: information quality, systems quality and service quality. Each one has to be measured separately, because singularly or jointly, they will affect subsequent system usage and user satisfaction.

In 1996, Wang and Strong proposed their DQ framework [35], which will be discussed in more detail in the following section. In their framework they categorised characteristics/attributes in to four main types/factors: intrinsic, accessibility, contextual and representational. This method of categorising IQ factors and attributes proved to be a valuable methodology for defining IQ [33]. Lately, several quality management projects in business and government have successfully used this framework [33, 43-45].

After Wang & Strong DQ framework, diverse research efforts were spent in order to identify IQ dimensions in deferent contexts. Although these frameworks varied in their approach and application, they shared some of the same characteristics concerning their classifications of the dimensions of quality [46].

In 1996, Gretz focused on finding possible solutions for the problems regarding modelling and managing data quality and integrity of integrated data [47]. He proposed a taxonomy of data quality characteristics that includes important attributes such as timeliness and completeness of local information sources. While Redman's [48] work aimed to set up practical guidelines to analyze and improve information quality within business processes, he proposed a number of quality attributes grouped into six categories: Privacy, Content, Quality of Values, Presentation, Improvement and Commitment. In the same year, Zeist & Hendricks [49] identified 32 IQ sub-characteristics grouped in 6 main IQ characteristics which covered functionality, reliability, efficiency, usability, maintainability and portability.

Unlike general purpose IQ framework, in 1997 Jarke and Vassiliou [50] proposed a special purpose framework where he used the same hierarchical design established by Wang & Strong. He defined IQ criteria depending on the context and requirements for specific application; Data Warehouse Quality (DWQ). In his framework, they linked each operational quality goals for data warehouses to the criteria which describe this goal. The main defined criteria are accessibility, interpretability, usefulness, believability, and validation

In 1998, Chen [51] gave a list of IQ criteria with no special taxonomy. He, however, proposed a goal-oriented framework focusing mainly on time-oriented criteria such as response time and network delay. One year later, Alexander & Tate [52] proposed their framework for IQ in web environment. This framework consisted of 6 main criteria; authority, accuracy, objectivity, currency, orientation and navigation. In the same year, Katerattanakul & Siau [44] adapted Wang & Strong DQ framework to propose their four categories IQ framework of individual websites. Furthermore, Shanks & Corbitt [53] recommended a semiotic-based quality framework for information on the Web. This framework includes four semiotic levels. Syntactic level to insure that information is consistent whiles the Semantic level focuses on the information completion and accuracy. Pragmatic level is the third level which covers the usability and the usefulness of the information. The forth level is the social level ensures information understandability. Within their framework there are 11 quality dimension distributed within the identified levels.

Dedeke in 2000 [54] developed a conceptual IS quality framework that includes 5 categories; ergonomic, accessible, transactional, contextual and representational quality. Each category consists of number of quality dimensions such as; availability, relevancy and conciseness. Whilst Zhu & Gauch [55] described 6 quality metrics for information retrieval on the web; these are availability, authority, currency, information-to-noise ratio and cohesiveness.

Leung adapted Zeist & Hendricks's quality framework in 2001 [56] and applied it to Intranet applications. He defined 6 main IQ characteristics; functionality, reliability, usability, efficiency, maintainability and portability. Each quality characteristic in the proposed framework includes numbers of sub-characteristics.

Several research in IS quality were undertaken in the following year, Eppler & Muenzenmayer [57] suggested two main manifestations for their proposed framework; content quality and media quality. The content quality is focused on the quality of the presented information and it consists of two categories; relevant information and sound information. Whereas media quality is focused on the quality of the medium used to deliver the information and it includes optimized process category and reliable infrastructure category. Each category in the framework contains number of quality dimensions. Khan [58] categorised IQ depending on the context of the system. The framework divided IQ into two main quality types; product and service quality. Moreover, it divided these two types into 4 quality classifications and each classification into number of quality dimensions. The quality classifications are sound information, useful information, dependable information and usable information.

In addition, Klein [45] conducted a research in the same year to identify five IQ dimensions chosen from Wang & Strong's DQ framework to measure IQ in Web context; accuracy, completeness, relevance, timeliness and amount of data. Mecella [59] also proposed an initial framework for quality management in Cooperative Information System (CIS). This framework includes a model for quality data exported by cooperating organizations and the design of an infrastructure service and improving quality.

More recent, in 2005 Liu and Han [60] mentioned 6 key dimensions for IQ; source (focused on information availability), content (focused on information completeness), format and presentation (focused on information consistency), currency (focused on information currency and timeliness), accuracy (focused on information accuracy and reliability) and speed (focused on how easily information is downloadable).

Besiki et all in 2007 [61] introduced a general framework for IQ assessment. This framework consists of a comprehensive taxonomy of IQ dimensions, and provides a straightforward and powerful predictive method to study IQ problems and reason through them in a systematic and meaningful way.

Lately, Kimberly et all presented in 2009 [62] a model for how to think about IQ depending on the application context; they identified number of common IQ metrics. Kargar & Azimzadeh [63] also presented an original experimental framework for ranking IQ on the web log. The results of their research revealed 7 IQ dimensions for IQ in web log. For each quality dimension, quality variables associated coefficients were calculated and used so that the proposed framework is able to automatically assess IQ of web logs. In the same year Thi & Helfert [64] conducted a research aimed to propose a quality framework based on IS architecture. In their research they identified quality factors for different construct levels of IS architecture. Moreover, they also presented impacts amongst different quality factors which help to analyze the cause of IS defects.

In this part we gave a brief review of the researches conducted to formulate (data/information) quality frameworks within information systems. However in the next section we will focus on Wang and Strong's DQ framework as we will use it as a base for this research to measure IQ in e-learning systems along the dimensions of the framework.

2.3.4 Wang and Strong's Data Quality Framework

Wang & Strong's DQ framework, one of the most comprehensive, popular, remarkable and cited DQ frameworks, was established by Richard Wang and Diana Strong in 1996 [35]. Their framework was designed empirically by asking users to give their viewpoints about the relevance of the IQ dimensions to capture the most important aspects of DQ to the data consumer. Their hierarchical conceptual framework of DQ is shown in Figure 5

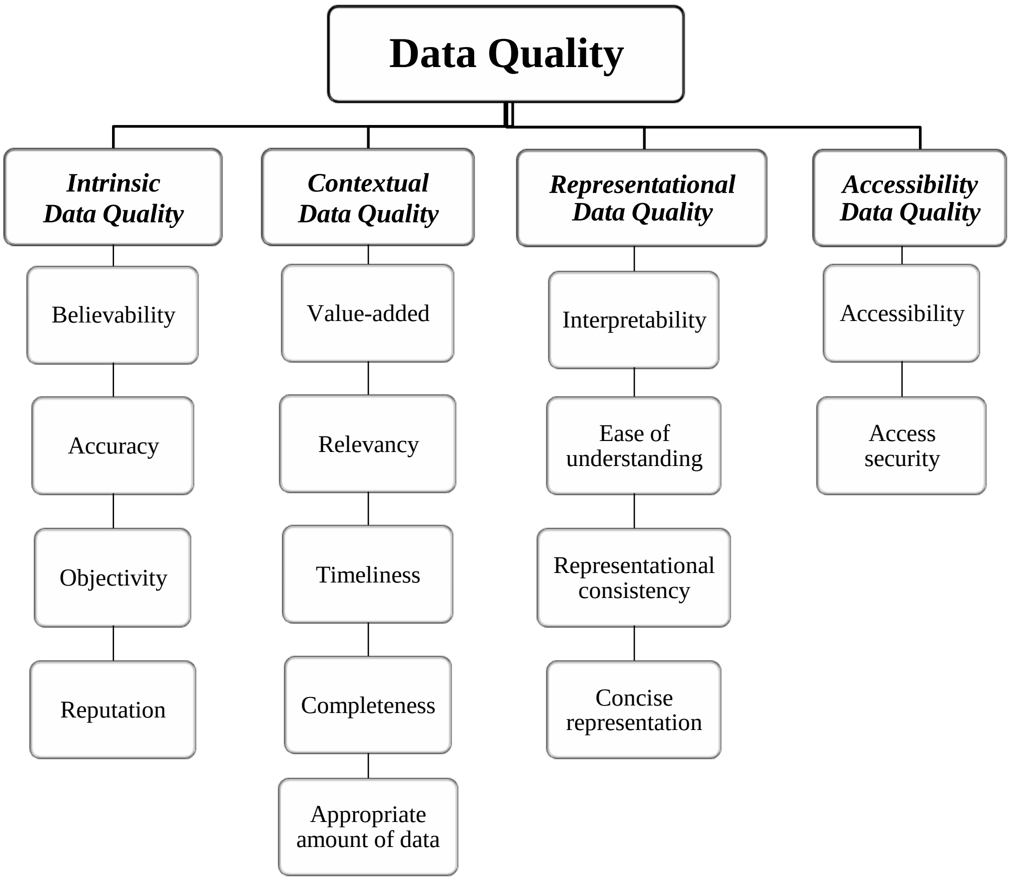


Figure 5. Wang and Strong data quality framework. Source[35 p.21] (reproduced by kind permission of the author)

In their framework, Wang and Strong classified quality dimensions into four groups:

- ***Intrinsic DQ:*** refers to the quality dimensions originating from the data on its own. This aspect of quality is independent of the user's perspective and context.
- ***Contextual DQ:*** focuses on the aspect of IQ within the context of the task at hand. In this group, the quality dimensions are subjective preferences of the user.

Contrary to the first group, DQ dimensions cannot be assessed without considering the user's viewpoint about their use of information.

- ***Representational DQ:*** is related to the representation of information within the systems.
- ***Accessibility DQ:*** refers to the quality aspects concerned with accessing distributed information.

The defining feature of this particular study is that quality attributes of data were collected from the data consumer instead of being defined theoretically or being based on the researchers' own experiences. Their research can provide a basis for measuring DQ/IQ along the dimensions of this framework.

2.3.5 Summary

In this part of the literature review we shed some light on the use of the terms "data quality" and "information quality", we also discussed the concept of IQ within information systems and considered how it could be defined. We also gave a historical review of the researches conducted to formulate (data/information) quality frameworks within information systems, focusing on Wang and Strong's DQ framework which will provide a good basis for this research to measure IQ in e-learning systems along the dimensions of this framework.

However, this research will also investigate the possibility of integrating a web mining approach, a data gathering technique, in order to automate the evaluation process. It seems logical, therefore, that the available methods for web mining and information extraction are now reviewed. These will be discussed in the next section.

2.4 Information Extraction and Web Mining

This study focuses not only on the evaluation of IQ in the context of e-learning systems, but also it will investigate the possibility of integrating a web mining approach, a information extraction technique, in order to automate the evaluation process. This part of the literature review will provide a brief overview of the information representation on the web. It will also focus on web mining definitions and categories, and the idea of information extraction.

2.4.1 Information on the Web

Today, the web is becoming more popular and interactive information publishing mediums and the levels of web information are growing rapidly. Moreover, the web holds a huge amount of distributed information for news, education, government, e-commerce and various other information services [65]. Also, the web contains a rich and dynamic collection of information about hyperlinks, webpage access and usage [66]. In fact, today web users can access vast amounts of information, however, it becomes ever more difficult to weed out the irrelevant and discover the relevant which has drawn attention to a fundamental issue: information overload [67].

The nature of web information is unstructured, thus it can only be understood by humans, but the massive amount of available information means that it can only be processed efficiently by machines [68]. A lack of metadata, data about data, represents another challenge when dealing with the published information [69].

To be able to cope with these challenges researchers started to apply techniques from data mining and machine learning to web data and documents [70]. Web mining applications help users in finding, sorting and filtering the available information, while the semantic web aims to make the data machine understandable as well [71].

2.4.2 Web Mining

Extracting useful or valuable information from the web is usually referred to as "web mining" [72]. It refers to the application of data mining methods for the discovery of useful information on the web [73].

In the literature, several definitions exist relating to web mining. It could be generally defined as the automated discovery and analysis of useful information published in web documents and services using data mining methods. It is a large and new area converging from several research districts, such as database, information extraction and artificial intelligence [70]. Web mining techniques could be used to solve the information overload problem [74].

2.4.2.1 Web Mining Categories

There are three categories for web mining according to the different sources of the target data [72]:

- ***Web content mining:*** which addresses the discovery of knowledge from the content of web pages, thus, it includes the target data contained in a web page as text, images, multimedia, etc.
- ***Web usage mining:*** which addresses the discovery of knowledge from user navigation data while surfing the web, thus, this includes the target data contained in users' log files.
- ***Web structure mining:*** this addresses the discovery of knowledge from hyperlinks on the web.

This broadly used categorisation of web mining started in 1997 when Cooley, Srivastava and Mobasher introduced web content mining and web usage mining [73], while web structure mining was added in by Kosala and Blockeel in 2000 [71].

The focus in this research will be on web content mining as a technique to automate the extraction process of the information needed in the quality measurement.

2.4.2.2 Web Mining and Information Extraction

Natural language (NL) texts are used mostly as digital information storage mediums. The main goal of information extraction (IE) is to find the required information in NL texts and store this information in a way that is suitable for automatic querying and processing. IE involves defining output representations or templates and searching only for information that fits the defined representations [75].

2.4.3 Summary

Within this section of the literature review a brief idea of information representation on the web was provided. It also shed some light on the web mining definition and considered the categories of web mining, finally, the idea of information extraction was noted.

2.5 Conclusions

The literature review provided a general background to the subject of e-learning, including the definitions, types and the concepts of quality, IQ within information systems, and web mining as an information extracting technique. The literature offered here mainly focused on the sub-topics of the larger research areas which will be directly applicable to this research.

The present study differs from earlier studies in many aspects, as it tackles the following gap in knowledge:

- Since the specified context and the perspectives of the users need to be considered when defining quality in any IS, and because the majority of the revised framework for quality in IS are considering IQ in general context, the main consideration for this study is identifying the main quality standards in e-learning context from the users' perspective.
- Most of traditional methodologies which are used to evaluate the quality of the distributed learning materials are qualitative and are generally user oriented. The main problem with these measurement methods is that they can be uncertain and inaccurate, and also require large amounts of effort and time. Consequently, we aim to take advantage of web mining techniques to automate the retrieval process of the information needed in the quality measurement.

Chapter 3: Contributions towards an Information Quality Framework in an E-learning Context

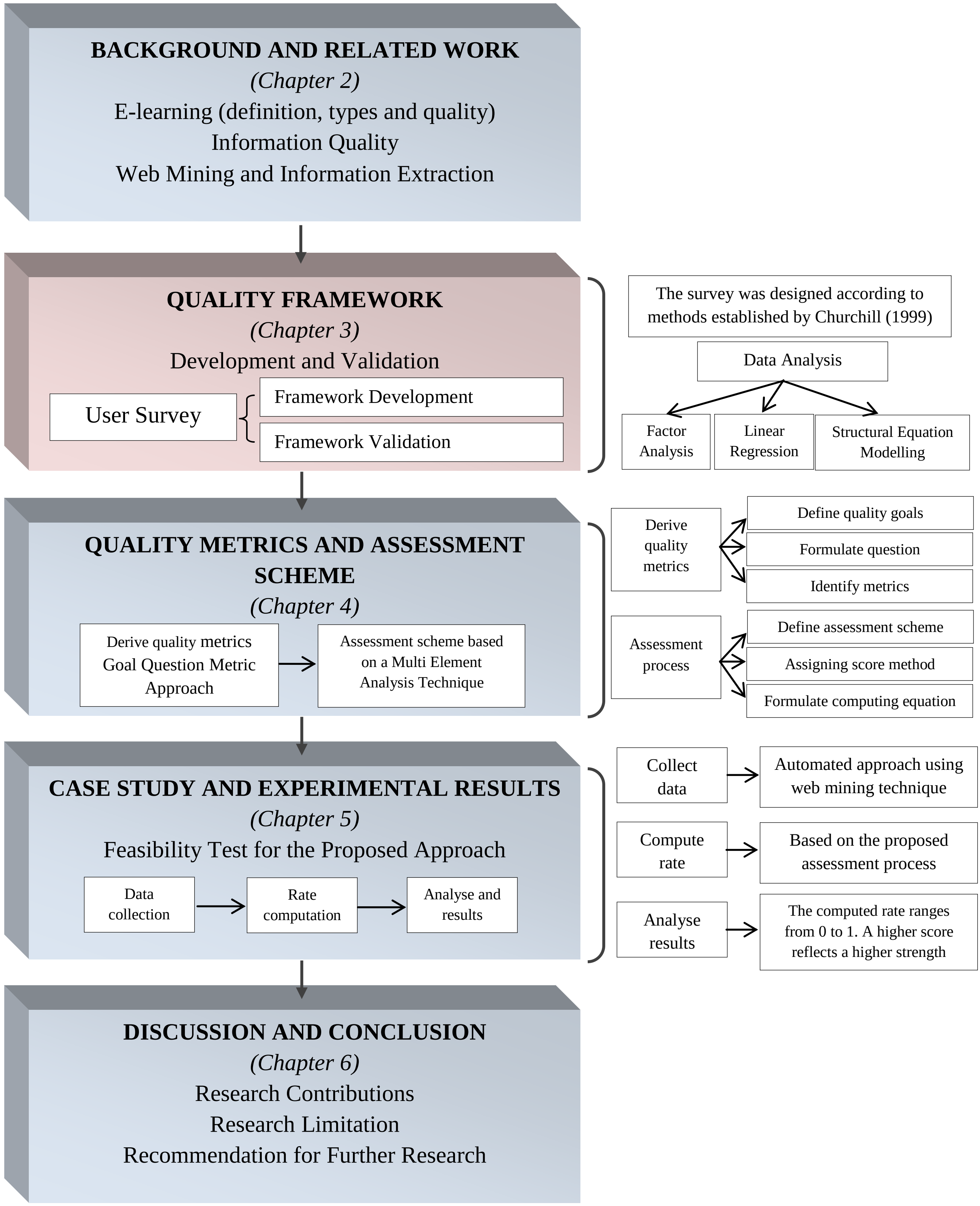


Figure 6. Thesis structure / Chapter 3

3.1 Introduction

This chapter focuses on concepts of IQ in the context of e-learning systems, particularly on identifying the key attributes for IQ from the users' perspective, in order to build a quality framework to measure the quality of the content provided by e-learning systems. It is essential to identify quality attributes accurately, as they provide the building blocks for further research into the quality of e-learning systems and for IS in general [2]. In our study, Wang and Strong's DQ framework [35] was extended and used as a reference point, owing to its popularity and acceptance by the IS quality community.

This chapter will focus on the development process for an IQ Framework within an e-learning context. Also, it will present an empirical validation of the proposed framework.

3.2 An Extended Information Quality Framework

We started the framework development by adopting Wang and Strong's DQ framework and using it as a guideline, as it is generally accepted by the IS quality community [35]. This framework was chosen owing to its popularity as it proved to be useful for many research regarding IQ in information systems. Moreover, as mentioned in the chapter 2, the framework was designed empirically by asking the users and it focus on DQ/IQ which could form a appropriate basis for this research. More details about this framework can be found in the literature review (section 2.3.4).

Although their quality model provides a good base for our research in measuring IQ in e-learning systems due to the attributes of the framework, it was necessary to extend it to include any undiscovered quality attributes that may have arisen in recently published research in the area of the quality in information systems. After Wang and Strong's DQ framework was identified, diverse research efforts were spent in order to identify IQ attributes in different contexts, as mentioned previously (section 2.3.3).

We extended Wang and Strong's DQ framework by examining seventeen frameworks within the recently published literature, covering the timeline from 1996 to 2007 just before starting the next step of the research [44, 45, 47-61]. Interestingly, it was found that there was no general agreement on the used attributes. We, however, decided to use only the quality attributes which were mentioned by at least two authors within the examined frameworks. In general, we found nineteen common quality attributes used in most of the frameworks. Fifteen of these were already used in Wang and Strong's framework. Table 2 gives the frequency of the appearances for every attribute in the examined frameworks, while Table 3 summarises the occurrences of these attributes within the examined frameworks.

Quality attributes	Frequency
Accuracy	15
Believability	7
Consistency	8
Objectivity	8
Reputation	4
Appropriate amount of data	10
Completeness	13
Relevancy	12
Timeliness	14
Value-added	2
Verifiability	3
Concise representation	3
Ease of understanding	5
Interpretability	2
Representational consistency	7
Accessibility	11
Access security	5
Availability	3
Response time	8

Table 2. Attributes frequencies in the examined frameworks

Information quality		Information quality frameworks																
Quality factors	**Quality attributes**	Gertz & Managing [47]	Redman [48]	Zeist & Hendriks [49]	Jarke & Vassiliou [50]	Chen et al [51]	Alexander & Tate [52]	Katerattanakul & Siau [44]	Shanks & Corbitt [53]	Dedeke [54]	Zhu & Gauch [55]	Leung [56]	Eppler & Muenzenmayer [57]	Kahn et al [58]	Klein [45]	Mecella [59]	Liu & Han [60]	Besiki et al [61]
Intrinsic Factor	Accuracy	√	√	√	√	√	√	√	√	√		√	√		√	√	√	√
	Believability				√		√	√	√		√		√	√				
	Consistency				√				√	√			√	√		√	√	√
	Objectivity		√				√		√		√	√	√	√	√			
	Reputation						√	√	√								√	
Contextual Factor	Appropriate amount of data	√	√	√		√	√			√		√	√	√	√			
	Completeness	√	√		√	√			√	√		√	√	√	√	√	√	√
	Relevancy		√	√	√	√	√	√		√	√	√			√		√	√
	Timeliness	√		√	√	√	√		√	√	√	√	√	√	√	√	√	
	Value-added									√				√				
	Verifiability		√										√					√
Representational Factor	Concise representation		√						√				√					
	Ease of understanding			√					√	√			√	√				
	Interpretability		√		√													
	Representational consistency		√		√		√	√					√	√			√	
Accessibility Factor	Accessibility			√			√	√	√	√	√	√	√	√			√	√
	Access security			√	√							√	√					√
	Availability	√	√		√													
	Response time				√	√				√	√	√	√	√			√	

Table 3. Comparison between the occurrences of quality attributes in different iq frameworks

These attributes are grouped into four main factors, as defined within the Wang and Strong framework. The nineteen initial quality attributes, which were identified in the examined frameworks, will be used as an extended framework and, therefore, as a fundamental base through which to discover the important quality attributes from an e-learning user's perspective.

3.2.1 Proposal for a Preliminary Extended Framework

Our proposal to update Wang and Strong's DQ framework initially comprised four further quality attributes: consistency, verifiability, response time and availability. Therefore, the extended framework consists of four quality factors and nineteen quality attributes, as shown in Figure 7.

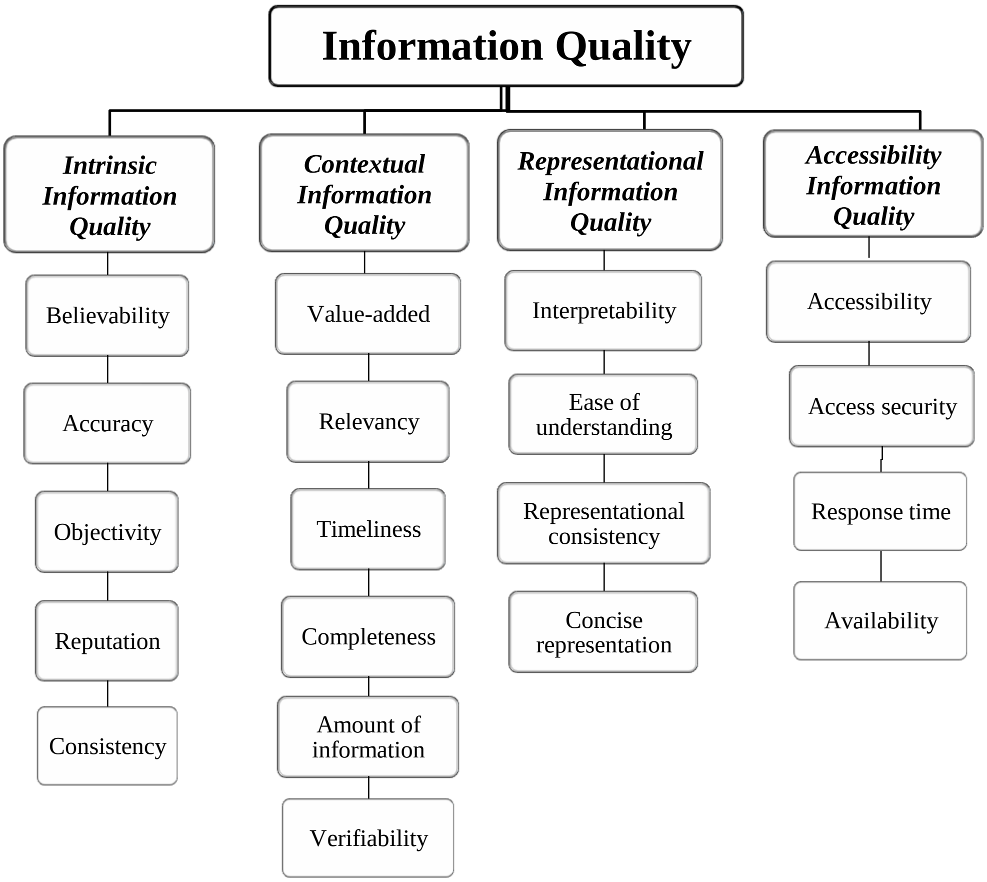


Figure 7. Proposal for an Extended Framework

3.2.2 Questionnaire to Determine Users' View of the Relative Importance of Information Quality

Although quality frameworks help in the measurement procedure, defining quality using a framework is not enough because, as mentioned before, IQ is dependent on the application context [3] For that reason, the identified quality attributes were arranged in a questionnaire format to determine a user view of the relative importance of quality attributes in an e-learning system. This questionnaire[2] sought to gather the views of end-users with regard to the importance of IQ attributes in e-learning systems. It also sought to give an indication about the importance and relevancy of these quality attributes for the users. This would help in ranking these attributes in order to develop an IQ framework and quality metrics to measure the quality of information provided by e-learning systems. In order to prevent response bias in the questionnaire it is important for each question to be clear and precise[76].

The questionnaire in this study was a cross-section survey [76], and we designed it according to the methods introduced by Churchill in 1999 [6], where he distinguished between three types of research design methods depending on the research questions and aims; these are descriptive research, causal research and exploratory research, which is the case in our study. For this kind of study, the most appropriate sampling method is probability sampling, where each respondent is chosen randomly from the population [77]. The questionnaire was performed on a sample from a population of persons involved in academic work and dealing with e-learning systems on a regular basis. Respondents included both learners and teachers. The questionnaire was distributed to the respondents via e-mail because of its reduced cost, decreased transfer time and its convenience for respondents. *Surveymethods.com*, an online survey

[2] The survey is illustrated in Appendix I.

software application, was used to create the survey, deploy it via e-mail, and collect -0 respondent data through its graphical based analysis module. The questionnaire was planned to take less than five minutes to complete. It consisted of three parts:

Part 1. Obtained a brief profile of the respondent.

Part 2. Addressed the user's attitude and usage of the Internet in general and e-learning systems specifically.

Part 3. Asked respondents to rank the nineteen quality attributes depending on their importance. A five-point Likert scale, ranging from (1) 'Very Important' to (5) 'Not at all Important' was used to measure respondents' perspectives [78, 79].

As suggested by Gillham and Oppenheim [80, 81], a pilot experiment was conducted on a representative sample of five individuals that were randomly selected and questionnaire statements were modified based on the results of this initial experiment.

Responses were collected from 315 e-learning system users[3], from 24 different countries as illustrated in Figure 8.

46% of the respondents were from Saudi Arabia, 26% from the United Kingdom, 12% from Romania and the rest of the respondents were from 21 remaining countries.

Moreover, 57% of the participants were female, and 43% were male. All respondents in the sample were e-learning users from different learning institutes. Of the respondents that contributed, the majority, 66%, used e-learning as students, 29% were teachers and/or authors of the learning materials, while 5% used e-learning systems for other purposes, such as librarians and technicians as shown in Figure 9.

[3] As recorded on 5th of March 2009.

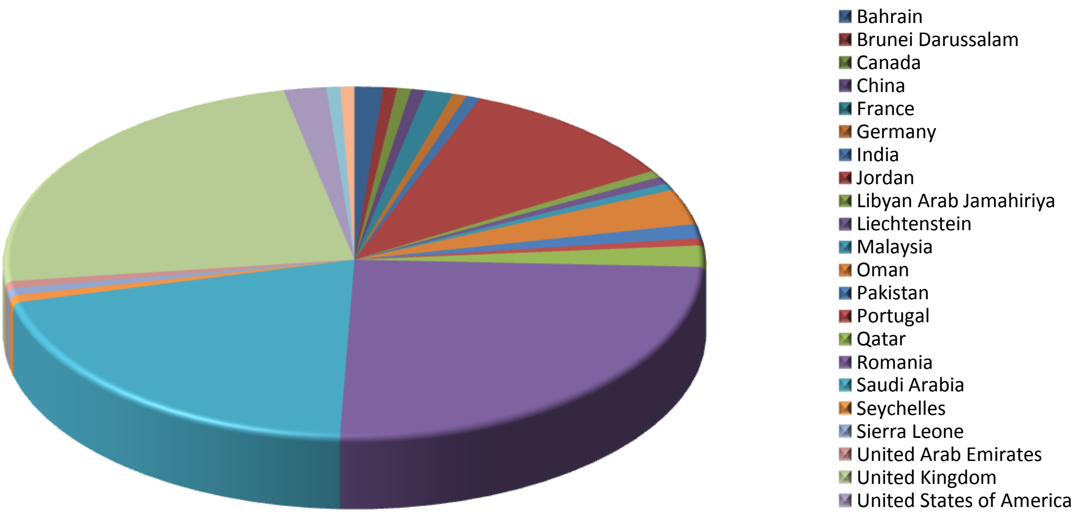


Figure 8. Distribution of the collected sample within the 24 countries

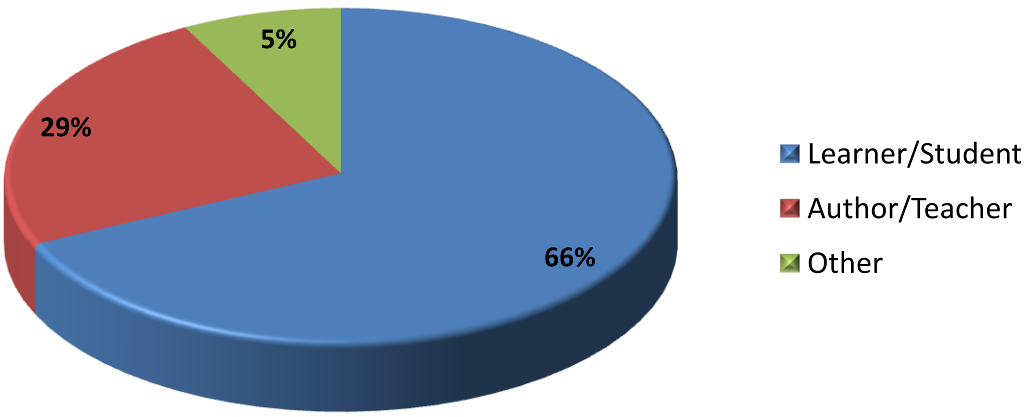


Figure 9. Distribution of e-learning users within the collected sample

In addition, participants held various qualifications: 40% held Bachelor's degrees, 33% had Master's degrees, and 20% had a PhD, while the remaining 7% held those listed as 'others'. However, as recommended by Bloom [82] students educational background could influence their answers, it is suggested that more studies be undertaken to examine the influence of these differences on the users' perspective of the quality.

3.2.3 Data Analysis

The collected data from the third part of the questionnaire was analysed using SPSS, to identify the most important quality attributes in the area of e-learning systems and to build the final quality framework. The steps of data analysis are illustrated in Figure 10.

First, a frequency analysis was conducted for each variable (attribute) to check for major mistakes and missing data. The results for the attribute frequency analysis in each factor showed that the data was valid and ready to be analysed.

Then a reliability test was conducted, which is generally used to indicate the level to which research results would be the same if the investigation was to be repeated with a different sample or at a later date. In this study the most accepted test of inter-item consistency reliability was used; that is, the Cronbach's coefficient alpha [83, 84]. Based on Sekaran [85], reliabilities less than 0.6 were considered to be poor, those in the 0.7 range were acceptable, and those over 0.8 were good: the closer to 1.0, the better the reliability coefficient. It is generally agreed that the minimum acceptable value of Cronbach's alpha is 0.70 [86, 87], but this could be reduced to 0.6 for exploratory research [88]. The Cronbach's alpha values for the attributes in each quality factor gave an acceptable reliability level with 0.712, 0.735, 0.781, and 0.625 for intrinsic, contextual, representational and accessibility information quality, respectively.

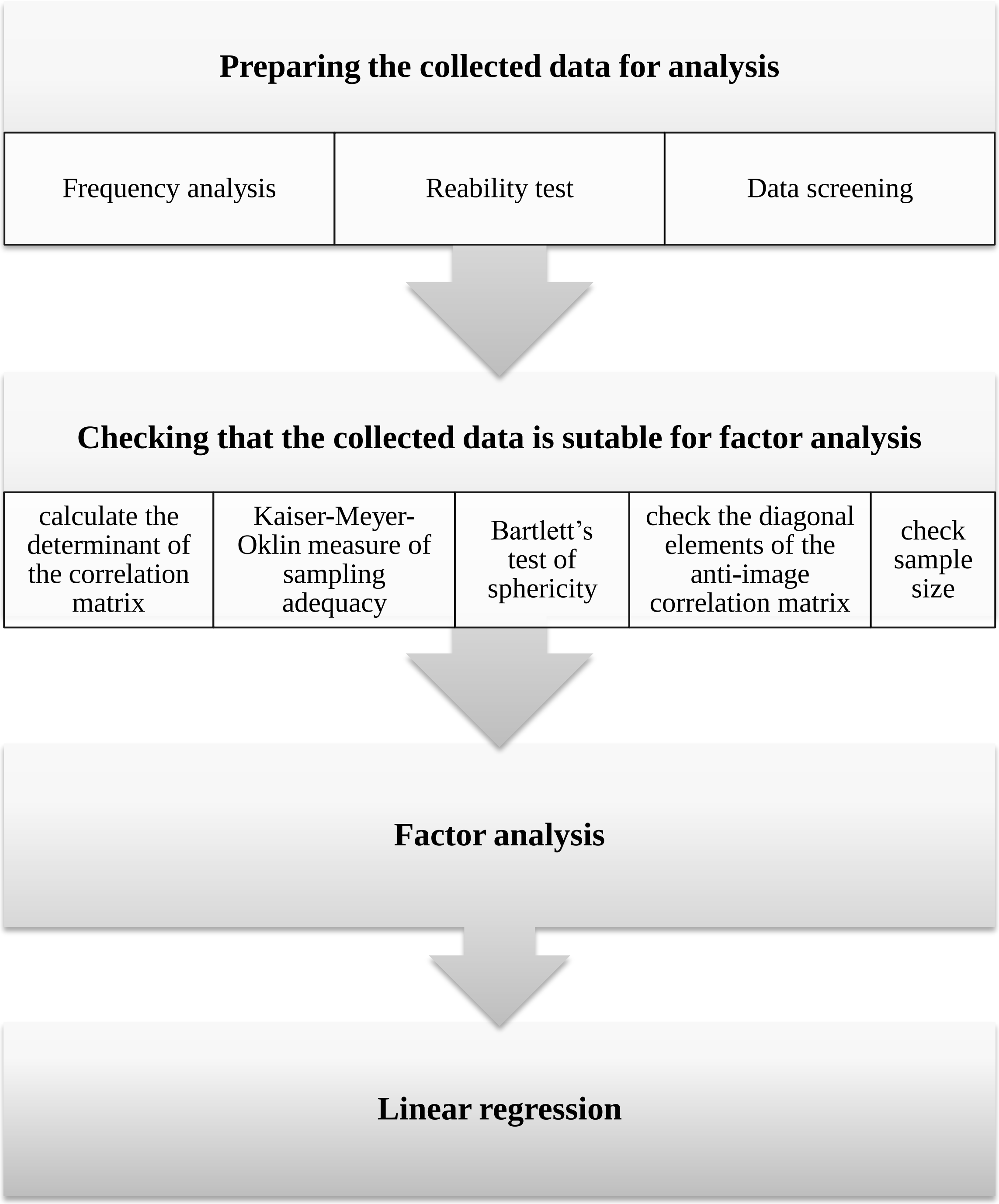


Figure 10. Steps for data analysis

However, in order to increase the reliability levels we decided to carry out data screening, in response to Churchill's recommendation in [89]. So, the collected data was screened by discarding items that showed very small corrected item-total correlations (<0.40). Because of this test, timeliness and value-added attributes were deleted from the contextual factor, and the access security variable was deleted from the accessibility factor, which left only 16 attributes in the framework. As a result, the reliability coefficient increased to 0.712, 0.748, 0.781, 0.668 for intrinsic, contextual, representational and accessibility factors, respectively.

The next stage was to conduct a factor analysis procedure with *varimax* rotation to check the dimensionality of the construct. To choose the cut-off value, there is no fixed measure. It depends on the purpose of the study at hand. Haire [90] recommended that item loadings >0.30 are considered significant, >0.40 are more important, and >0.50 are considered very significant. While the aim of this study is to recognise the most important and significant quality attributes, it was decided to use a cut-off point of 0.50 for item loadings and an eigenvalue of 1.

Before conducting the factor analysis and to make sure that the collected data was suitable for this kind of analysis, the determinant of the correlation matrix was calculated: it was 0.002, which was greater than the necessary value of 0.00001. As a result, we were confident that multicollinearity would not cause any problems during the analysis [91].

Moreover, the Kaiser-Meyer-Oklin (KMO) measure of sampling adequacy and Bartlett's test of sphericity, which is illustrated in Table 4, were performed.

The KMO statistic is a value between 0 and 1; a value close to 1 indicates that patterns of correlation are fairly compact and as a result factor analysis should give distinct and reliable factors [91]. Values between 0.5 and 0.7 are average, values between 0.7 and 0.8 are good, values between 0.8 and 0.9 are great and values above 0.9 are excellent [92].

Kaiser-Meyer-Oklin Measure of Sampling Adequacy		.879
Bartlett's Test of Sphericity	*Approx. Chi-Square*	1845.750
	Df	120.000
	Sig.	.000

Table 4. KMO and Bartlett's test

In addition, the significance value for the Bartlet's test should be less than 0.05 [91]. In this study, the KMO value was 0.879, which is in the range of 'great' and the Bartlett's test was highly significant. Therefore, we should be confident that factor analysis is appropriate for this data.

In addition to examining the overall KMO statistics, it was essential to check the diagonal elements of the anti-image correlation matrix[4] (which illustrates the KMO value for individual attributes); as in the overall KMO value, these values have to be greater than 0.50 for all attributes [91]. For this data set, the values exceeded the requirements and were in the range between 0.828 and 0.934.

With regard to the sample size, Comrey [93] and Lee stated that 300 is a good sample size for factor analysis, 100 is poor while 1,000 is excellent. As the number of our sample exceeded 300 respondents, it can be accepted that the sample size is appropriate for this type of analysis.

As a result of conducting the factor analysis, we obtained the eigenvalues associated with each factor before extraction, after extraction and after rotation; these are listed in Table 5.

[4] See Appendix II.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulativ e %
1	6.058	37.865	37.865	6.058	37.865	37.865	3.735	23.343	23.343
2	1.401	8.753	46.619	1.401	8.753	46.619	2.627	16.418	39.762
3	1.188	7.424	54.043	1.188	7.424	54.043	2.285	14.281	54.043
4	.974	6.090	60.133						
5	.879	5.497	65.630						
6	.786	4.914	70.544						
7	.657	4.106	74.650						
8	.598	3.740	78.390						
9	.565	3.530	81.920						
10	.556	3.477	85.397						
11	.532	3.324	88.721						
12	.455	2.843	91.565						
13	.412	2.576	94.141						
14	.366	2.286	96.426						
15	.306	1.911	98.338						
16	.266	1.662	100.000						

*Extraction Method: Principal Component Analysis.

Table 5. Total variance explained

Before extraction, SPSS identified 16 factors within the data set. SPSS then extracted all factors with eigenvalues greater than 1, which left us with three factors.

From the scree *plot* shown in Figure 11, it can be seen that the point of inflexion on the curve on three factors is in conformity with the results shown in Table 5. Thus, the most suitable way was to stick with three factors.

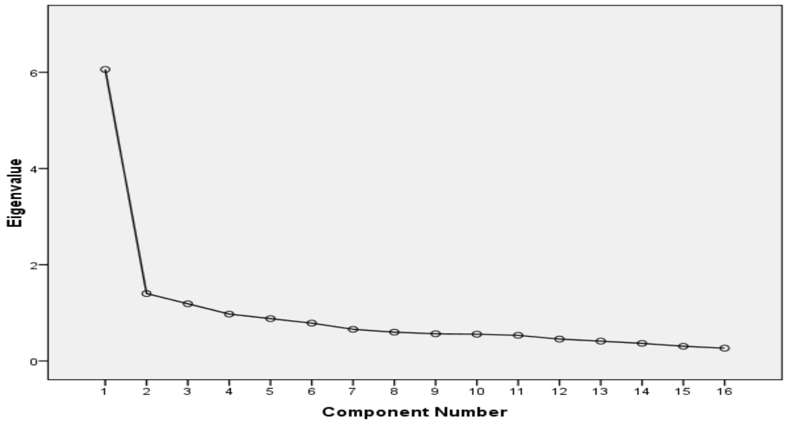


Figure 11. Scree plot for point of inflexion for each attribute

The matrix of the factor loadings for each attribute on each factor 'the rotated component matrix' is shown in Table 6.

Quality attributes	Component		
	1	2	3
Believability			.689
Accuracy			.736
Objectivity			.765
Reputation	.673		
Consistency			
Relevancy		.541	
Completeness	.582		
Amount of information	.607		
Verifiability	.695		
Interpretability			
Understandability	.643		
Representational consistency	.596		
Conciseness	.809		
Accessibility		.667	
Response time		.623	
Availability		.782	

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 6. Rotated Component Matrix factor loadings for each attribute

Factor loadings less than 0.5 were displayed because we asked for these lodgings to be suppressed. As a result, the suppressed attributes, consistency and interpretability, were discarded which left only 14 attributes in total.

Analysis findings show that there are three information quality factors in e-learning systems not four, as proposed previously. It was recognised that contextual and representational quality factors measure the same aspects from an e-learning system user's perspective. Therefore, a new quality framework, with 14 attributes of IQ in e-learning systems is proposed in order to measure three quality factors: intrinsic, contextual representation and accessibility IQ. The new proposed framework is shown in Figure 12.

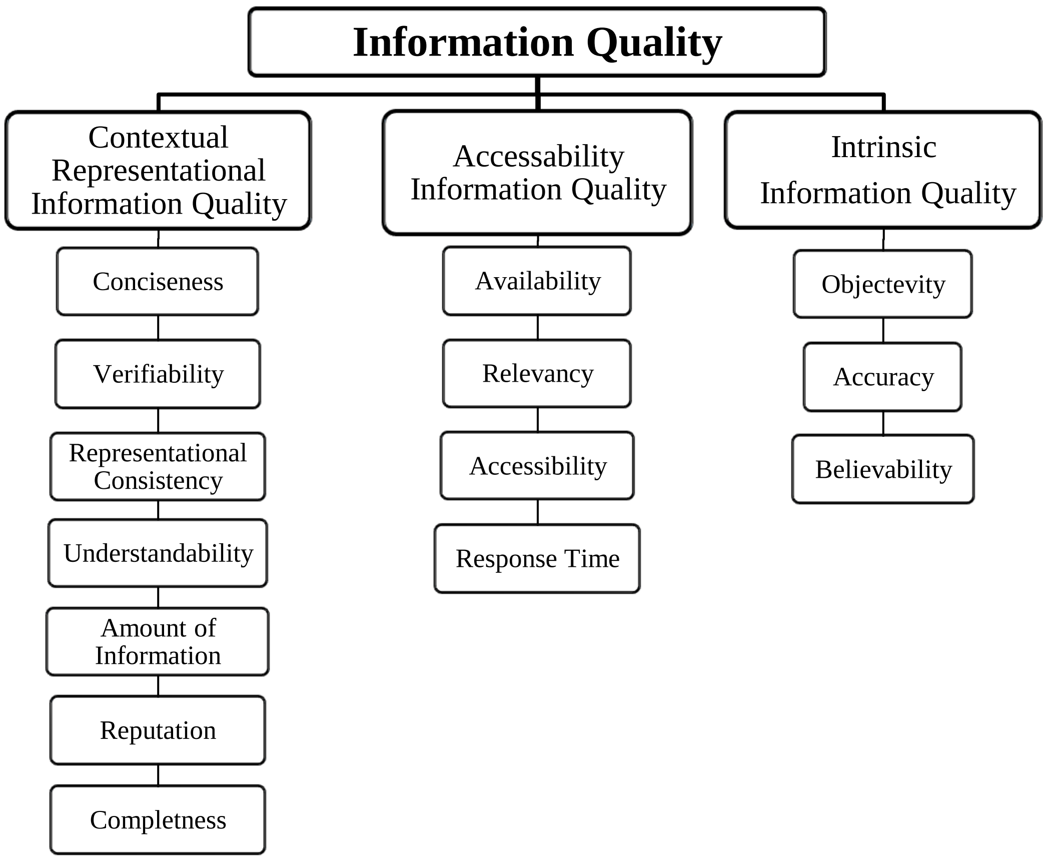


Figure 12. The new proposed framework for information quality in e-learning

Cronbach's alpha values were then calculated for the attributes in each new factor, which gave a good reliability level with 0.842, 0.697, and 0.665, for intrinsic, contextual representation and accessibility information quality, respectively.

Linear regression then was used to predict the factor scores from the attributes. Correlation coefficients can be obtained by squaring the "*partial correlation*" provided by SPSS which is "β" in the equation (1) which was used to calculate the relative importance of each variable in the correlated factor below[5].

$$relative\ importance\ for\ vi = \frac{\beta i^2}{\sum_i \beta i^2} \quad , \mathrm{i} = 1, \dots, 14 \qquad (1)$$

where βi was the partial correlation for the variable vi in the corresponding factor.

For example, for completeness in the first factor, it was 0.156^2 = 2.434%. These statistics will sum to less than 100%. To get them to total 100%, we divided each by the sum of all. The same logic was conducted to define the relative importance for each factor in the overall quality.

The zero-order correlations[6] are the loadings. One could define the relative importance of a variable as the amount by which the explained variance in the factor is reduced if the variable is removed from the regression model. That statistic measure is the squared semi-partial.

[5] According to Professor Karl L. Wuensch from East Carolina University and Dr. Andy field from Sussex University, (personal communication, November 2008)

[6] See Appendix III

3.2.4 Proposal for a Final Framework

The revised framework, after calculating the relative importance for each attribute inside the three quality factors and the relative importance for each factor in the overall quality, is proposed in Figure 13.

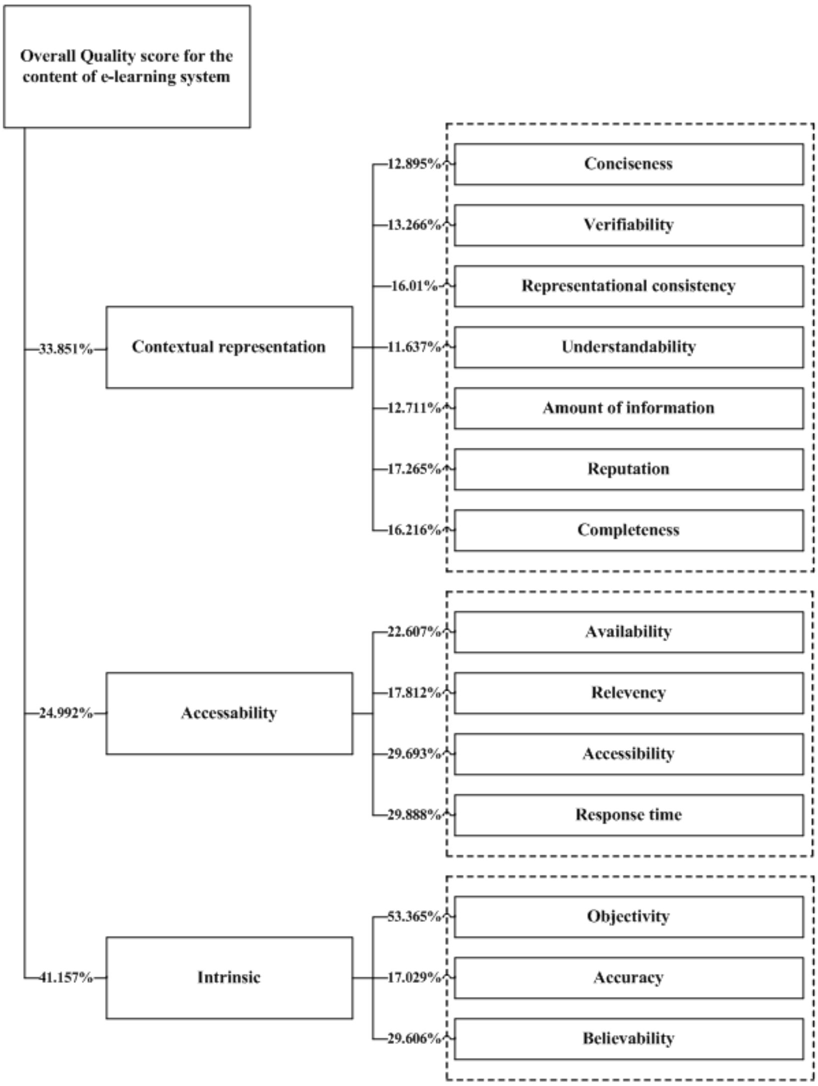


Figure 13. The proposal for the final framework of information quality in e-learning

The final framework consists of 14 quality attributes grouped in three quality factors: intrinsic, contextual representation and accessibility. The results show that the most important factor is intrinsic IQ with a relative importance score of 41.157% of the overall quality, while contextual representation and accessibility scored 33.851% and 24.992%, respectively. Objectivity is the most important dimension in the intrinsic factor. Reputation scored the highest relative importance within the contextual representation factor. Whereas accessibility and response time have almost the same relative importance within the accessibility factor with the scores 29.693% and 29.888%, respectively.

3.2.5 Summary

Based on the original questionnaire data and factor analysis, this part of the thesis presented a new quality framework to measure the quality of the content provided by e-learning systems. Moreover, linear regression was used to calculate the relative importance of each quality attribute within the main quality factors, and the relative importance of each factor in the overall quality. The next section aims to validate the quality framework using an empirical approach.

3.3 Gender and Cultural Differences in Users' Perceptions of Information Quality in E-learning Systems

When talking about the quality in e-learning, we should take into account the fact that the use of the Internet as a distribution channel for e-learning may be affected by the global nature of the Internet, therefore, we should not forget that learners work in national contexts, and have differences in other aspects, such as gender and cultural identity. Moreover, the widespread use of the Internet and technology all over the world has raised key questions about the relationship between cultures and technologies [94]. In this context, there are a large number of cross-cultural studies of people's use and perceptions of technology [94-100]. We will not go deep in summarizing this great body of research, but findings seem to suggest that, males and females might have different perceptions regarding the technology. Also, cultural differences may affect users' perceptions and uses of the technology and the Internet. For that reason we will give an example of how could the proposed framework be useful to examine the differences in users' perceptions of information quality in e-learning systems from a gender and cultural perspective.

Within our proposed quality framework we assigned a relative importance weight for each attribute within the main quality factors, and a relative importance weight for each factor in the overall quality score. The idea of using relative importance as a parameter for the measurement is important, since it provides the framework with the flexibility to be adopted and used in different e-learning environments and with different users. This flexibility allows overcoming of the problem of the users' differences.

This section aims to shed some light on gender and cultural related differences in user perceptions of the relative importance of the main quality factors within the proposed framework. Also, it examines whether the observed differences are significant enough to be considered. The approach taken in this section is to compare user perspectives in two different geographical cultures (Middle East and Europe). Besides that, we will examine the differences between male and female perceptions of information quality in e-learning systems. We will examine two main hypotheses: first, males and females will have different perceptions of information quality in e-learning, and second, that in a comparison of Middle Eastern and European e-learning system users, there will be cultural differences in the same subject as suggested in the related literature [94-100].

3.3.1 Outline of Participants

We used the data collected from the information quality framework development phase. However, we used the answers collected from the respondents without considering the educational background for the users which could be a rich area for more extensive comparison studies.

As mentioned earlier, the participants for this questionnaire consisted of a total of 315 e-learning system users, from 24 different countries. Most of the respondents (55%) were from the Middle East or Europe (40%) while the remaining 5% were from other countries, such as Malaysia, Canada or the USA. From the collected sample, Saudis constituted 83% of Middle Easterners, while 65% of Europeans were from Britain, as illustrated in Figure 14 and Figure 15 perceptively.

Moreover, our sample consists of 57% females, while 43% were male. All respondents in the sample were e-learning users from different learning institutes.

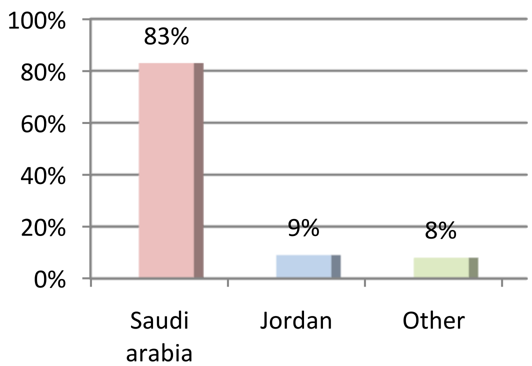


Figure 14. Distribution of the collected sample within the Middle East.

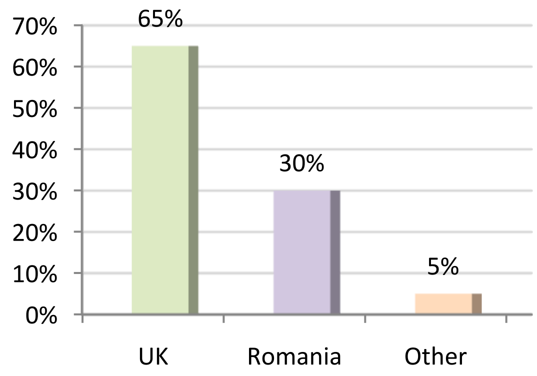


Figure 15. Distribution of the collected sample within the Middle East and Europe

3.3.2 Differences in Users' Perceptions

T-tests were applied to determine statistical significances between various measures in order to investigate differences in user perspectives of the relative importance of the main quality factors within the proposed framework, on the basis of gender and cultural background [101].

We analyzed the collected data from the third part of the questionnaire using SPSS; the results of the statistical tests are presented in this section of the thesis.

3.3.2.1 Gender Consideration

Table 7 presents the output for the t-test conducted to examine the differences in male and female perspectives of the relative importance of the three main quality factors. It contains the mean for male and female responses, mean differences between the two categories and the two-tailed p values for difference significance. Based on Field [102] and Babbie [103], any value less than 0.05 for the two-tailed p values indicates a significant statistical difference between the two samples .

	Mean		T	
	Female	Male	P=Sig. (2-tailed)	Mean Difference
Contextual representation	1.83	1.99	0.023*	-0.16
Accessibility	1.55	1.68	0.068**	-0.13
Intrinsic	1.69	1.55	0.012*	0.14

* p<0.05significant difference.
** p>0.05 No significant difference.

Table 7. T-test results for gender differences

Derived from the p values associated with each quality factor in table 7, which gives 0.023, 0.068 and 0.012, for contextual representation, intrinsic and accessibility of information quality, respectively, we can conclude that our first hypothesis, which stated that males and females will have different perceptions about information quality in e-learning, is supported by the achieved results. This is particularly evident for the first and second quality factors (contextual representation and intrinsic), where the values are less than 0.05. Consequently, we can assume that there will be a significant difference between female and male views regarding contextual representation and intrinsic information quality, while the differences regarding the accessibility factor are not significant enough to be taken into account.

Moreover, and looking at the mean differences in the same table, we can predict that females rate the contextual representation quality factor significantly higher in terms of its importance, while the intrinsic information quality factor seems to be more important from the male point of view [102]. However, the details of these differences will be examined when we calculate the relative importance of each quality factor of each category.

Also, linear regression was used to predict the relative importance of each factor for males and females; the results are shown in Figure 16. From the results demonstrated in this chart we can conclude that our assumptions regarding gender related differences were correct. The differences associated with the first two quality factors (contextual representation and intrinsic) were noticeable, while the difference associated with the third quality factor was negligible.

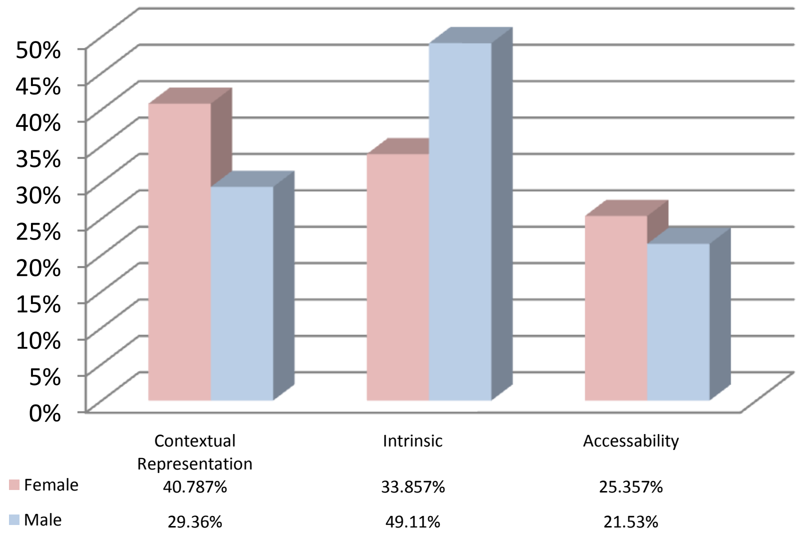


Figure 16. Female and male perspectives of the relative importance of the main quality factors

3.3.2.2 Cultural Consideration

Considering the p values calculated in table 8, which shows 0.54, 0.13 and 0.21, for contextual representation, intrinsic and accessibility factors of information quality, respectively, it is clear that the second hypothesis, which stated that Middle Eastern and European e-learning system users will have different perceptions of information quality in e-learning was not supported by the achieved results, where all values were more than 0.05.

	Mean		T	
	Middle Eastern	European	P=Sig. (2-tailed)	Mean Difference
Contextual representation	1.86	1.90	0.54*	-0.04
Accessibility	1.74	1.64	0.13*	0.09
Intrinsic	1.56	1.64	0.21*	-0.08

* p>0.05 No significant difference.

Table 8. T-test results for cultural background

Therefore, we can assume that there will not be significant cultural related differences in user perceptions of the relative importance of the main quality factors within the two different geographical cultures in the chosen sample [101]. More details about these differences will also be discussed when the relative importance for each quality factor will be presented.

Figure 17 illustrates Middle Eastern and European perspectives of the relative importance of the main quality factors. In fact, the recorded cultural related differences between e-learning system users' perspectives regarding information quality in e-learning in the Middle East and Europe were hardly noticeable.

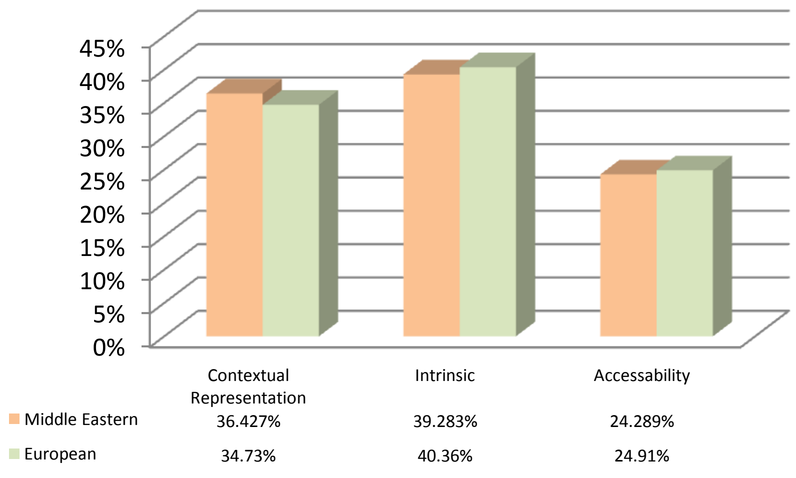


Figure 17. Middle Eastern and European perspectives of the relative importance of the main quality factors

3.3.3 Summary

This section showed e-learning system users had significant gender related differences in their perceptions of the relative importance of the main quality factors which were proposed in a previous work. Moreover, it showed that cultural related differences were not significant enough to be taken into account in the overall quality score.

Within this section we examined whether any significant gender and cultural related differences exist in user perceptions of the relative importance of the main quality factors within the proposed framework. In order to carry out this study, we benefited from the idea of using the users' relative importance of the quality factor as a parameter for the measurement within the proposed information quality framework. More investigations could be carried out in the future to examine whether the noticed gender differences remain constant across different cultures. Also, the same study could be applied using different cultural backgrounds.

3.4 Validation of the Framework

In this section, we present an empirical validation of the proposed framework. A Confirmatory Factor Analysis (CFA) approach was used to validate the proposed framework. CFA is widely used to determine the factorial construct validity for hypothesised models [7].

3.4.1 Validation Methodology

To perform the validation task, we decided to conduct a study by means of a survey. The 14 quality attributes, which we identify within our proposed framework, were used to build a five point Likert scale questionnaire [78], ranging from "Very Important" to

"Not at all Important". In addition to the IQ attribute items, the survey contained some demographic questions about gender, age and academic position. For each question, participants were asked to choose the answer that best described their level of agreement.

In order to reduce cost and decrease transfer time, the questionnaire was distributed to the respondents via e-mail. The collected data was then transferred into an SPSS file in preparation for analysis. The questionnaire was planned to take less than five minutes to complete.

3.4.2 Sample and Data Collection

In contrast to the development process, where the investigation used a cross-section survey performed on a sample chosen randomly (from a population of persons involved in academic work and dealing with e-learning systems on a regular basis) from 24 different countries, we found that a more focused and purpose-driven survey using a purposive sampling method was more appropriate for the validation procedure [77]. Therefore, in autumn 2009, we collected sample data from four academic institutions in Saudi Arabia. All of these selected institutions have implemented enterprise e-learning systems. Three of these institutions use Blackboard as a virtual learning environment (VLE) while the fourth institution uses WebCT. However, because some of these VLEs specifications could limit the contributions of e-learning developers which could influence users' answers, it is suggested that more studies be undertaken to examine the influence of these restrictions on the users' perspective of the quality.

A sample of 100 responses was obtained from e-learning users. It was decided that 100 participants would be acceptable as a minimum sample size for conducting a validation study using a confirmatory factor analysis (CFA) approach based on [7, 104].

Of those 100 e-learning users, 24 respondents were from teaching staff that use e-learning systems on a regular basis in the courses they teach. The 76 remaining respondents were students studying different computer science courses, with the age ranging from 18 to 22. Of the collected responses, 58 were from men while 42 of the respondents were female. All participating students were dealing with e-learning systems as part of their studies.

3.4.3 Data Analysis and Validation Results

The proposed model was validated using a CFA approach as a form of Structural Equation Modelling (SEM) goodness-of-fit test. SEM is a statistical methodology that takes a confirmatory and hypothesis-testing approach to the analysis of a structural theory bearing on some phenomenon [8]. CFA is a special case of SEM, also known as the covariance structure, which is widely used to test hypotheses about a particular factor structure. Moreover, CFA produces many goodness-of-fit measures to evaluate the model under validation [105].

The data was analysed using the Amos software package (version 16), which is a SEM software solution from SPSS Inc. A CFA approach was conducted using a maximum likelihood estimation to calculate the goodness-of-fit indices for the proposed framework.

The chi-squared divided by the degrees of freedom ($\chi 2/df$) statistic was the first fit index tested. A χ^2/df ratio < 2 indicates a good fit of the tested model to the empirical data.

We also tested the goodness-of-fit index (GFI) which is less than or equal to 1. A value of 1 indicates a perfect fit, so a value closer to 1 means a better fit for the model [106]. Other tested fit indices are the comparative fit index (CFI) and Tucker-Lewis index (TLI). CFI and TLI can be interpreted in a similar fashion to GFI. [107]. When GFI, CFI and TLI are greater than 0.9 the model may have a reasonably good fit [108].

One of the most informative fit indices, the root mean square error of approximation (RSMEA), takes into account the error of approximation in the population and the complexity of the tested model. RSMEA values less than 0.05 indicate a good model fit [8]. The goodness-of-fit indices are illustrated in Table 9, along with the recommended values.

Measure of fit	***Recommended values***	***Achieved values***
Chi-squared divided by degrees of freedom ($\chi2$/df)	Less than 2.0	1.06
Goodness of Fit Index (GFI)	Greater than 0.9	0.909
Tucker-Lewis Index (TLI)	Greater than 0.9	0.987
Comparative Fit Index (CFI)	Greater than 0.9	0.990
Root Mean Square Error of Approximation (RMSEA)	Less than 0.08	.024

Table 9. The goodness of fit indices

The $\chi2$/df, GFI, TLI, CFI and the RSMEA values indicate a good fit of the tested framework. It can, thus, be concluded that our proposed framework fits well with the empirical data.

The obtained results from this analysis support our validation process for the proposed framework. Moreover, it is possible to conclude, based on these results, that the proposed framework has been empirically validated and could be used to evaluate the quality of the information provided by distributed learning materials, from the users' perspective.

3.4.4 Summary

The main focus of this part was on the empirical validation of the proposed framework. The validation results are reported on the basis of data collected from four Saudi academic institutions. A confirmatory factor analysis (CFA) approach was used as a means of Structural Equation Modelling (SEM) goodness-of-fit tests to support our conclusions.

3.5 Conclusions

This chapter focused on identifying the key attributes for IQ from the users' perspective in order to build quality framework to measure the quality of the content provided by e-learning systems. The proposed framework consisted of 14 quality attributes grouped in three quality factors: intrinsic, contextual representation and accessibility. First, Wang & Strong's data quality framework was adopted and used as a reference point due to its popularity and acceptance by the information systems quality community. Seventeen frameworks were then reviewed from recently published literature to expand on Wang & Strong's framework; this included any undiscovered quality attributes. The identified quality attributes were arranged in a questionnaire format. Based on the collected data and factor analysis, a new quality framework was proposed to measure the content quality provided within an e-learning context. In addition, linear regression was used to calculate the relative importance weight for each factor and attribute in terms of the overall quality.

This framework could be used to provide a comprehensive indication of information quality in the context of e-learning systems. It could be useful to e-learning systems designers, providers and users as it provides a comprehensive indication of the quality of information in such systems. Moreover the idea of using relative importance as a parameter for the measurement is important, since it provides the framework with the flexibility to be adopted and used in different e-learning environments and with different users. This flexibility allows overcoming of the problem of the users' differences.

Moreover, in this chapter we presented an empirical validation of the proposed framework. Validation results were reported on the basis of data collected from an original questionnaire and a structural equation modelling (SEM) analysis, confirmatory factor analysis (CFA) in particular. The next stage will be the development of a set of quality metrics and an experiment to compute these metrics in chosen e-learning systems.

The next chapter will presents the metric identification process and uses a goal-question-metric (GQM) approach in order to quantify the quality of the information. It also discusses the definition of the assessment process, based on a multi element analysis technique.

Chapter 4: Contributions towards Information Quality Measurement Scheme and Metrics Definition

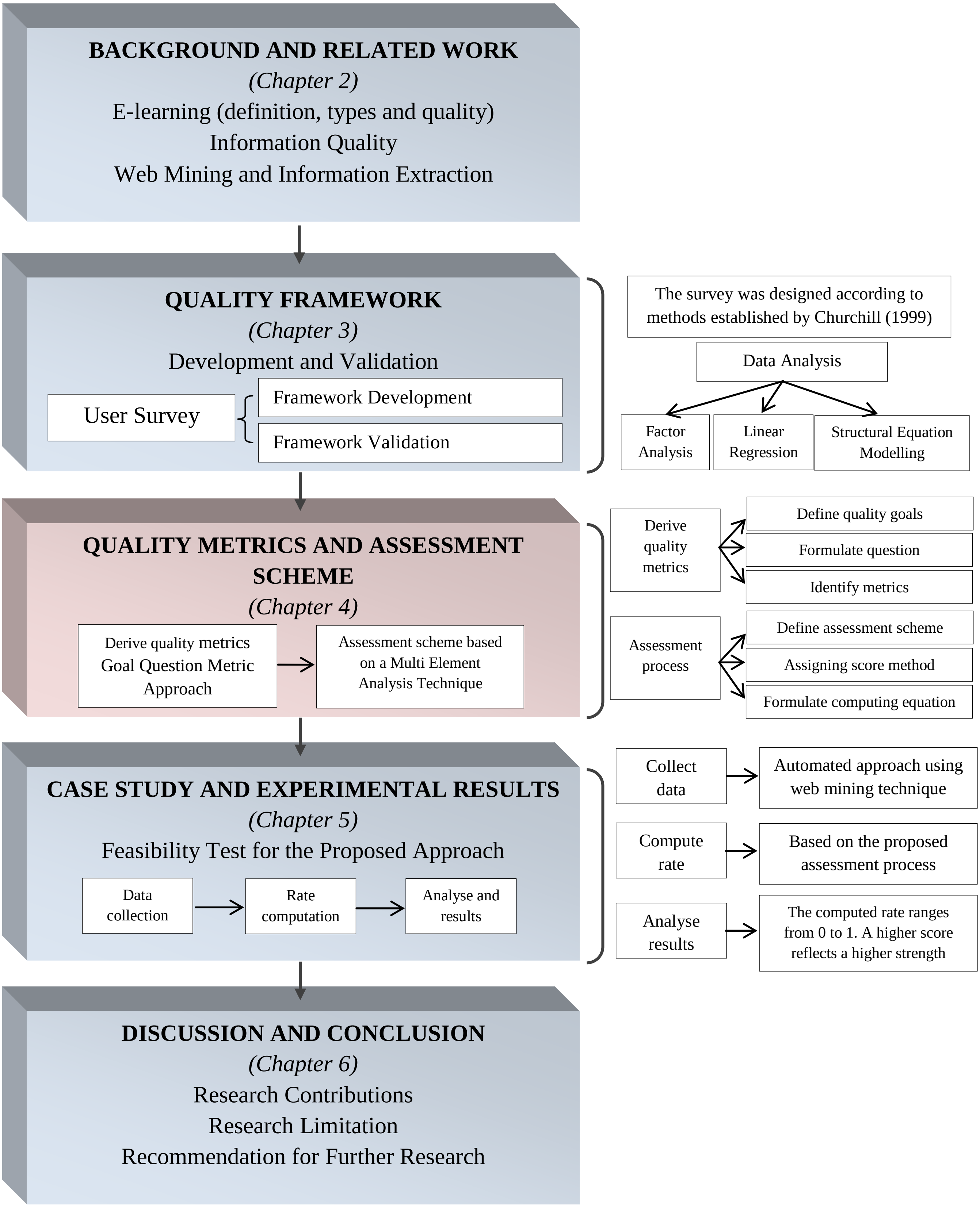


Figure 18. Thesis structure / Chapter 4

4.1 Introduction

Defining a suitable set of measurement metrics is considered to be one of the most important issues for any evaluation process, this enables the quantitative evaluation of the quality level, and supports the foundation for decision making [109].

This chapter focuses on the metric identification to quantify the quality of the information, in order to facilitate the evaluation, comparison, and analysis of IQ. A goal question metric (GQM) approach was used; a goal-oriented measurement strategy consists of deriving measures from measurement goals to ensure the consistency and completeness of measurement plans [110].

This chapter is organised as follows: the next section provides an overview of the GQM approach and discusses the application of the GQM approach in order to determine the quality metrics in an e-learning context. Then, a discussion about the definition of the assessment process, based on a multi-element analysis technique will be provided.

4.2 Measurement Metrics Identification for Information Quality

In order to identify suitable quality metrics for a specific domain, metric proposals should address users' needs in the domain context. To deal with this issue, the GQM paradigm is widely applied to define product quality metrics as a goal-oriented approach [56].

4.2.1 Goal Question Metrics Approach

This approach was originally proposed to evaluate defects in the NASA Goddard Space Flight Centre environment. It involved a number of case study experiments; its use has now been expanded to larger contexts to include different types of experimental approaches [111].

The GQM approach is based upon the assumption that to reach a purposeful measurement, first the project's goals should be specified. Thus, the final measurement can be analysed to determine whether the goals were actually achieved [112]. This method is useful in a goal-driven environment [9, 10, 109, 112-115].

The GQM model, shown in Figure 19, is a three level hierarchical structure:

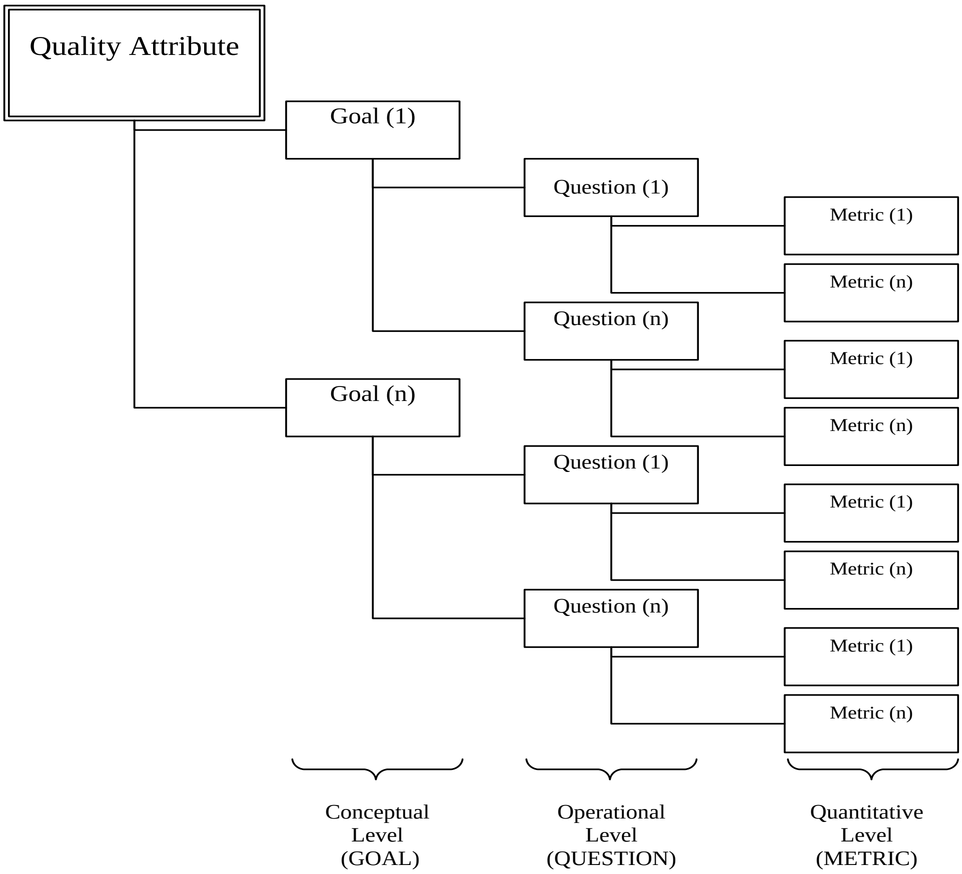


Figure 19. The hierarchical structure for the GQM model

- ***The first level:*** Conceptual level (GOAL): the major goals should be defined in this level.
- ***The middle level:*** Operational level (QUESTION): a set of questions should be formulated to be used for each goal; the answers to these questions could be used to determine whether the specified goals were met.
- ***The last level***: Quantitative level (METRIC): a measurement plan must be created detailing a set of measurement metrics associated with each question, thus providing quantitative answers.

By using the GQM approach, many current approaches to measurement are combined and generalised; these include processes and resources as well as product assessments. This approach is, therefore, flexible to be used in different environments, it has been applied in numerous organisations, including: NASA, Hewlett Packard, Motorola and Coopers & Lybrand [112].

Moreover, the GQM approach is considered suitable for application within web based systems where different types of web objects, such as java applets and scripts, are integrated to form a website [10].

4.2.2 Application of the Goal Question Metrics Approach

This section details the steps followed in applying the GQM approach to obtain the quality metrics to measure IQ within an e-learning context, which are based on the identified quality attributes in the proposed framework. The results of the obtained measurement metrics for IQ attributes in the intrinsic (InT), accessibility (AC), and contextual representation (CR) quality factors are presented in Figures 20, 21, and 22, respectively.

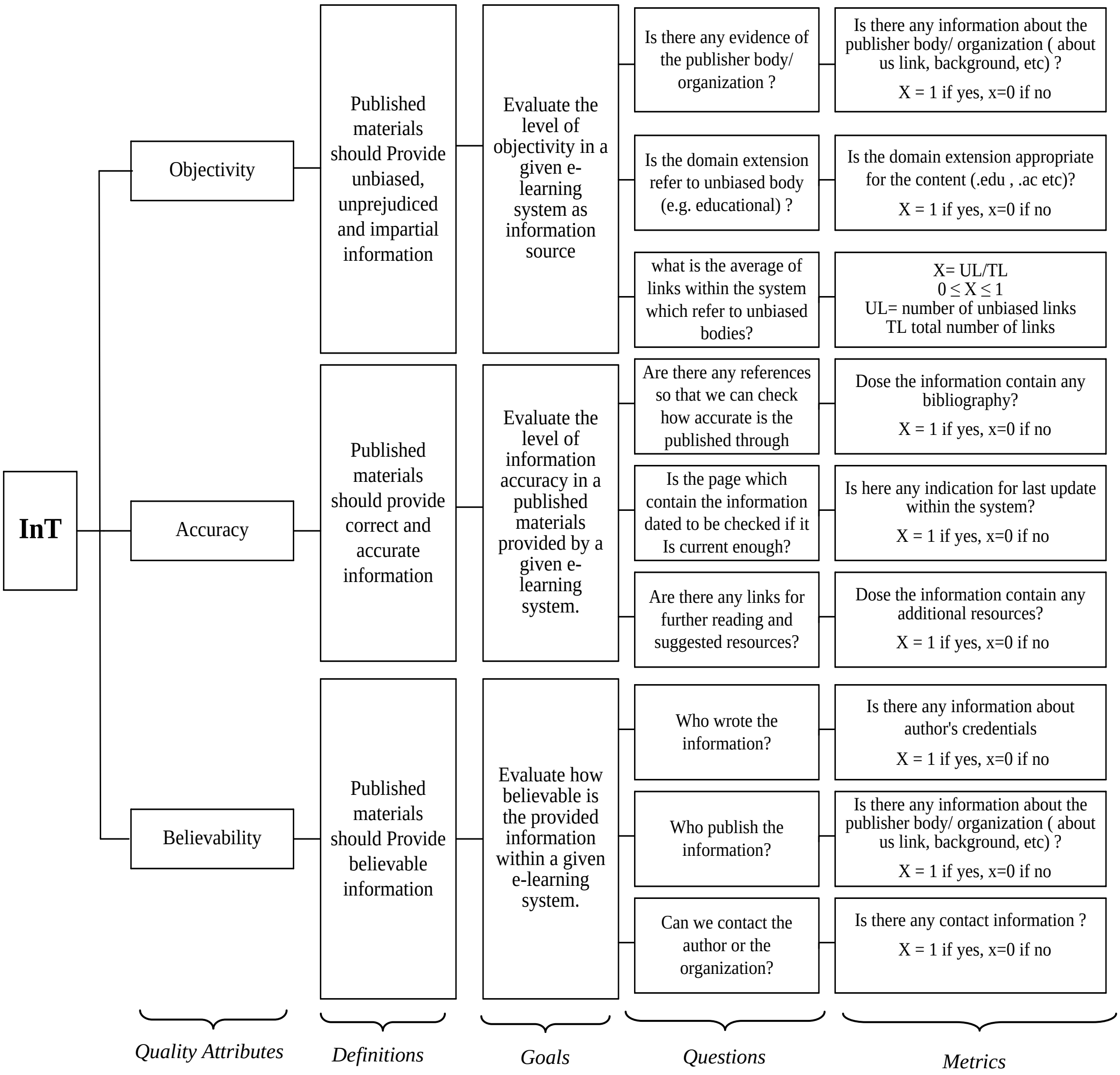


Figure 20. Quality metrics for IQ attributes within intrinsic quality factor

4.2.2.1 Quality Goals Definition

The starting point to creating the GQM plan is to identify quality goals, this step is very important and critical to ensuring the successful application of the GQM approach [115]. At the end of this phase, a set of goals associated with each quality attribute should be described.

In order to set the goal, five major elements should be identified: the object, purpose, quality focus, viewpoint and environment [116]. The example of goal setting for the "representation consistency" attribute can be represented as follows: evaluate the consistency of the information representation (object), for the purpose of quality measurement (purpose), with respect to information quality (quality focus), from the users' viewpoint (viewpoint), in a given e-learning system (environment). See Figure 22 for more details.

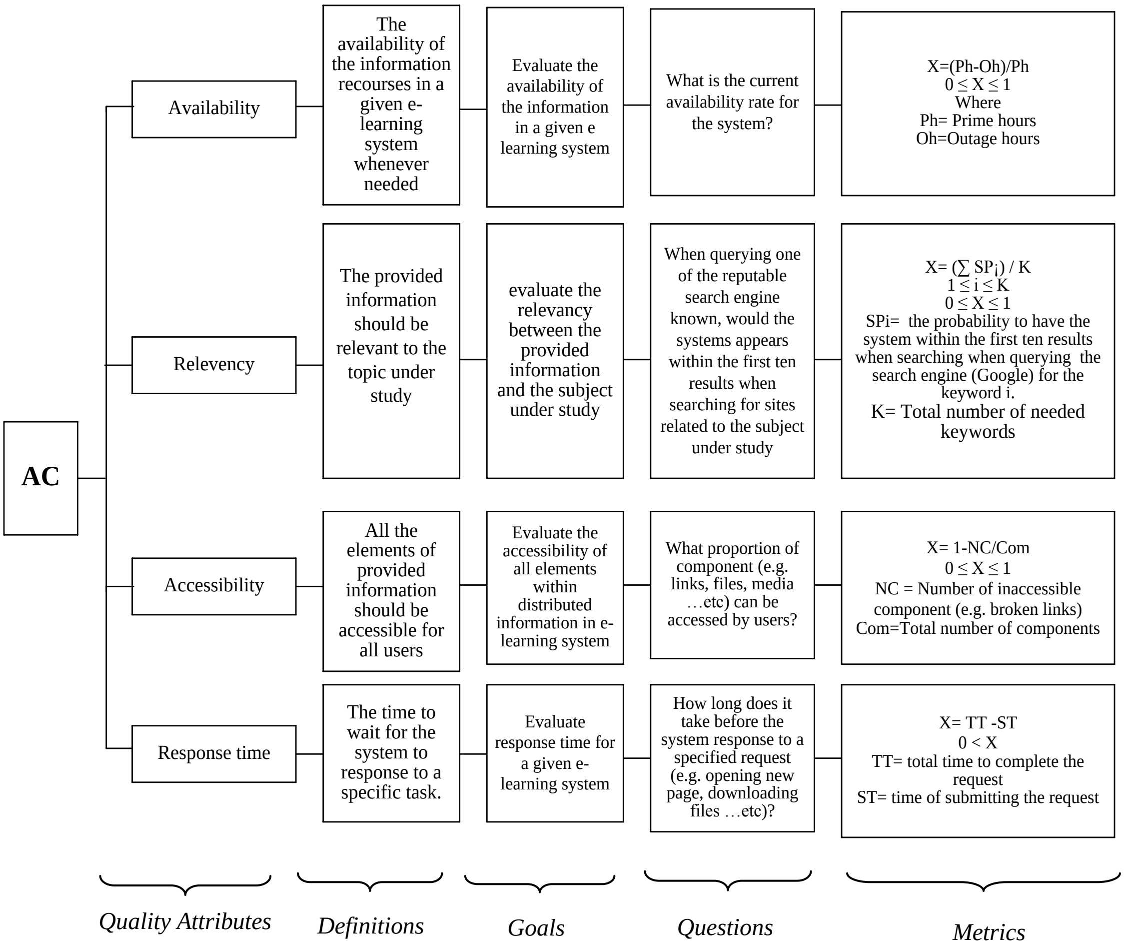


Figure 21. Quality metrics for IQ attributes within accessibility quality factor

4.2.2.1 Questions Formulation

The identified goals are then used to formulate relevant questions, to provide clearer definitions of the goals and to relate quality metrics in order to measure the quality in a quantifiable way. Questions should be defined in a manner that allows the answers to provide measurable values. To illustrate, it to focus on the goal mentioned previously could be represented as: "*How consistent is the representation of the provided information throughout the system?"* Figure 22 provides more details on formulating questions.

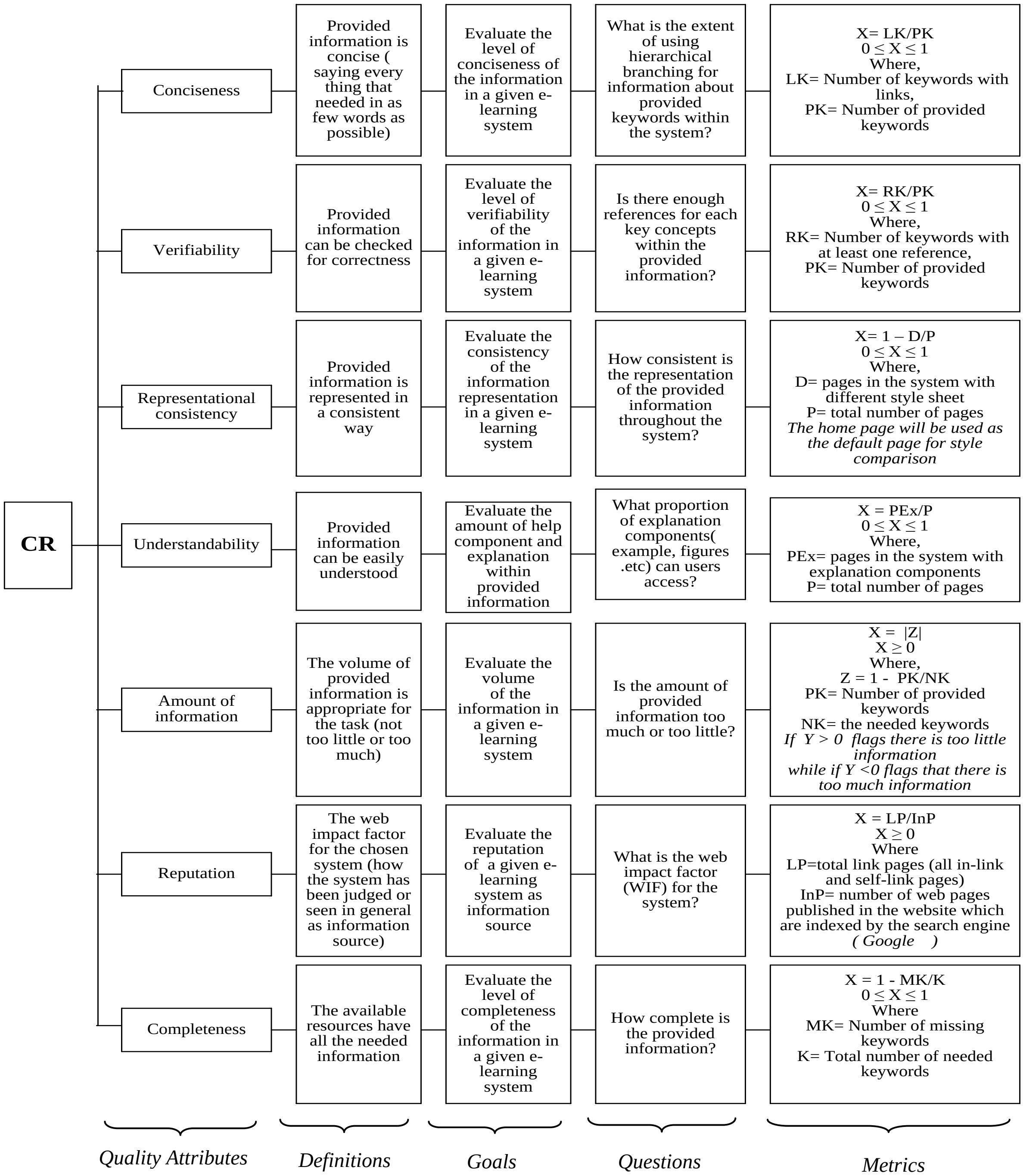


Figure 22. Quality metrics for IQ attributes within the contextual representation quality factor

4.2.2.2 Metrics Identification

It is anticipated that answers from the formulated questions will provide measurable metrics. The values derived from these metrics will be used later to calculate the quality score for each quality attribute, factor and for overall quality. The following is an example of the metrics associated with the “representation consistency” attribute (see Figure 22 for more details):

$$X=(1-D)/P$$
$$0 \leq X \leq 1$$

Where, D= pages in the system with different style sheet and P= total number of pages. The home page will be used as the default page for style comparison.

4.2.3 Summary

This section has discussed the approach used to develop quality metrics for measuring information quality in an e-learning context. The GQM approach was used as a mechanism for defining and interpreting operational and measurable quality metrics.

The next section will focus on defining the assessment scheme and process, based on a multi-element analysis technique.

4.3 Definition of the Measurement Scheme

The next step, after deriving the quality metrics, is to evaluate the identified metrics and use them to assess the IQ in an e-learning context. We will use a multi element analysis technique to reach an overall quality score for the provided information [117].

4.3.1 The Multi Element Analysis Technique

This technique was first proposed by the mathematician Zangerneister in 1970. Since then, the method has been successfully used as a system evaluation technique. Also, it was also engaged in the assessment of software maintenance tools [10]. Moreover, this technique was successfully used by Magnavox Electronic Systems Company (1990) in their evaluation of software development environments for Version 1 of the Advanced Field Artillery Tactical Data System [118].

To apply this method, three fundamental features should be presented: a hierarchical organisation of quality attributes/classes, a percentage weight for these attributes, and a numeric scoring for the final attributes/subclasses [119].

The method depends on a hierarchical organisation of quality attributes/classes and the percentage weights for these attributes. For each attribute/class throughout the construction, a percentage weight should be determined. The final attributes/classes are also assigned numeric scores to measure their performance.

The methodology for computing scores using this technique starts by assigning weights to all of the child nodes of each parent node in the attribute hierarchy. Then, scores should be assigned to the leaf nodes and the assigned weights are used to propagate quality scores to the root [118]. Intuitively, for all the child nodes of each parent node the sum of the weights should add up to 100 [120].

As the essential three components to apply this technique are already provided by our proposed framework, we decided to adopt this technique to define the measurement scheme and calculate the overall IQ score in an e-learning context.

4.3.1.1 *Measurement Scheme*

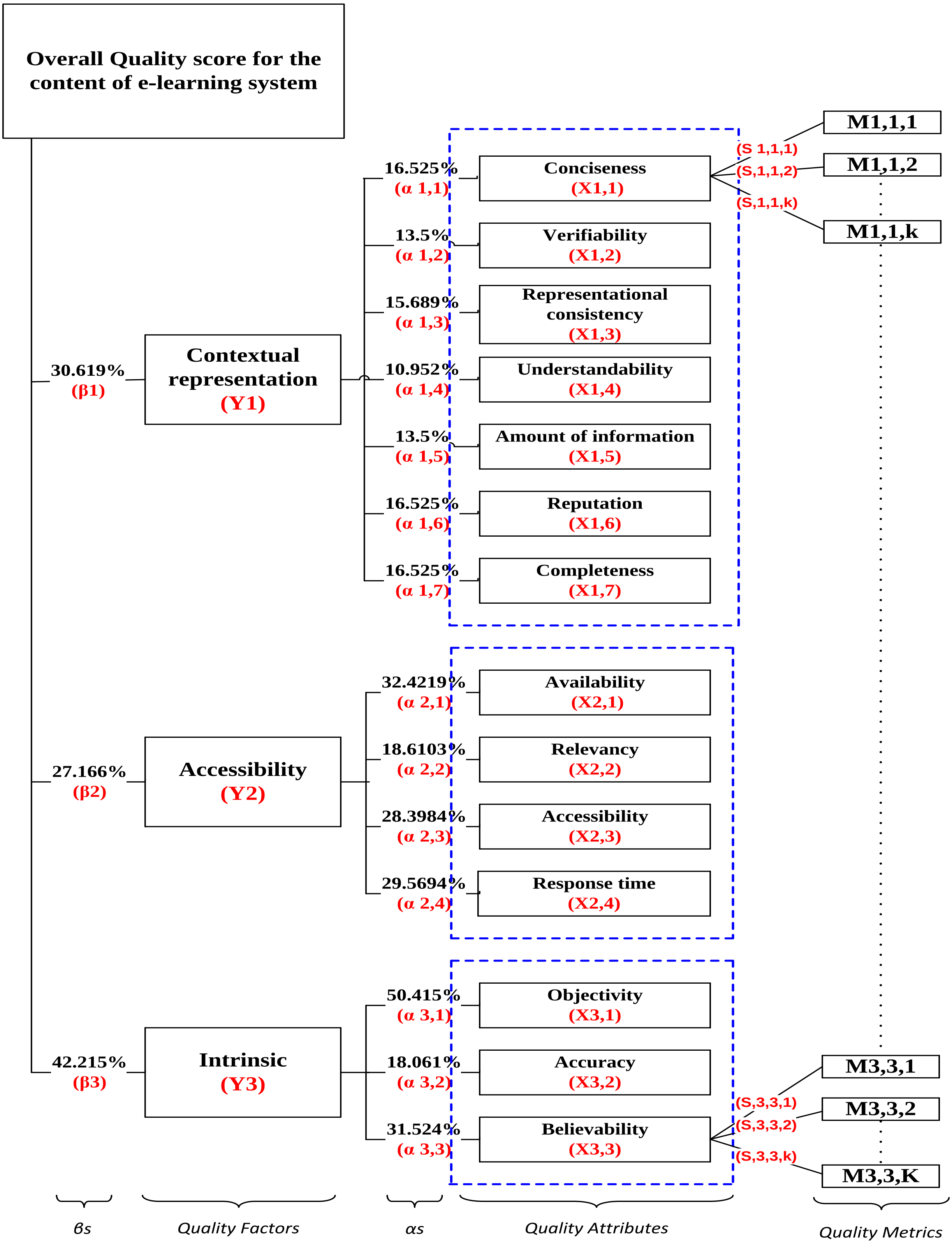


Figure 23. Structure of the measurement scheme

The idea of using the relative importance as a parameter for the measurement is important since it gives the framework the flexibility to be adopted and used in different e-learning environments and with different e-learning users, because the importance weights could be modified if a quality attribute or factor appears to no longer hold the same significance for the quality assessment.

In the proposed framework three main quality factors were identified and each factor consisted of a number of quality attributes (14 in total). We also assigned a relative importance weight for each attribute inside the main quality factors, and a relative importance weight for each factor in the overall quality score. The percentages of the weights of all attributes inside each factor add up to 100%. In the same way, accumulated percentage weighting of the quality factors would be always 100%. These identified quality factors, quality attributes and the associated metrics, which were developed in the last section, are mapped into a measurement scheme along with the assigned relative importance weights of quality factors and attributes.

Figure 23 shows the quality factors, attributes and the corresponding metrics along with the assigned relative importance. It also summarises the structure of the measurement scheme.

4.3.1.2 Score Assignment

Within each quality attribute, each defined metric will be assigned a direct score between 0 and 1; a normalisation method will be followed in cases where the values fall out of this range (as will be discussed later in section 4.4.1). For each metric, a higher value reflects a better quality score in terms of the corresponding attribute. The quality score for each attribute is calculated as the average of the values of the related metrics.

4.3.1.3 Rank Computation

To compute the overall quality score, this study will use the assigned relative importance weight for each attribute within the main quality factors, and a relative importance weight for each factor in the overall quality within the proposed framework. The following equations will be used for the calculation.

– ***To calculate the quality score for each quality attribute:***

$$X_{j,i} = AVG\left(S_{j,i,k}\right), k \geq 1 \qquad (2)$$

Where $S_{j,i,k}$ represents the score assigned to the quality metrics corresponding to the quality attribute i inside the main quality factor j.

– ***To calculate the quality score for each factor:***

$$Factor\ Quality\ Score(Y_j) = \sum_{i=1}^{m} \alpha_{j,i}\, X_{j,i} \qquad (3)$$

Where $\alpha_{j,i}$ represents the relative importance of the quality attribute i inside the main quality factor j, and $X_{j,i}$ represents the quality score given to the same attribute.

– ***To calculate the overall quality score:***

$$Overall\ Quality\ Score = \sum_{j=1}^{3} \beta_j\, Y_j \qquad (4)$$

Where β_j represents the relative importance of the factor j in the overall quality, and Y_j represents the quality score given to the same factor.

4.3.2 Summary

This section focused on defining the assessment scheme and process, based on a multi element analysis technique. A brief idea of the multi-element analysis technique was provided and was followed by a description of the measurement scheme, scoring method and computation of quality scores. However, before conducting the experiment there are two main issues to be considered; these will be highlighted in the next section.

4.4 Key Issues for Measurement Process

4.4.1 Range of Metrics Value

As mentioned earlier, for each defined quality metric there will be a corresponding direct score between 0 and 1. However, some metrics' values fall out of this range. Thus, the linear scaling transform method will be followed as a normalisation technique.

$$Y=(X-\text{min}[x_1,x_N]) / (\text{max}[x_1,x_N]-\text{min}[x_1,x_N]) \quad (5)$$

Where, min and max are the minimum and maximum values of variables.

4.4.2 Keywords Definition

As a number of the identified metrics depend somehow on the keywords for the published information, it seems logical to think of a suitable method to define and specify the needed keywords to be used in the identified metrics. There are several ways to do so; we could consult universities' publications and module specifications, look for the keywords within books in the associations in the topic, or talk to human resources professionals in the same field.

However, when conducting this experiment, we decided to consult publications in the same field and then ask five professional experts in the subject related to the information under assessment.

4.5 Conclusions

This chapter discussed the approach used to develop quality metrics for measuring IQ in an e-learning context. The GQM approach was used as a mechanism for defining and interpreting operational and measurable quality metrics. More explanation and examples for the defined metrics will be provided in the next chapter.

This chapter also focused on the development of the assessment scheme using a multi element analysis technique.

The next chapter will detail the technical experiment, where web mining will be used as a data collection technique in order to automate the assessment process. By analysing the collected data, quality scores and feedback about the IQ, within a given e-learning system, suggestions for future improvements and recommendations can be provided.

Chapter 5: Case Study and Experimental Results

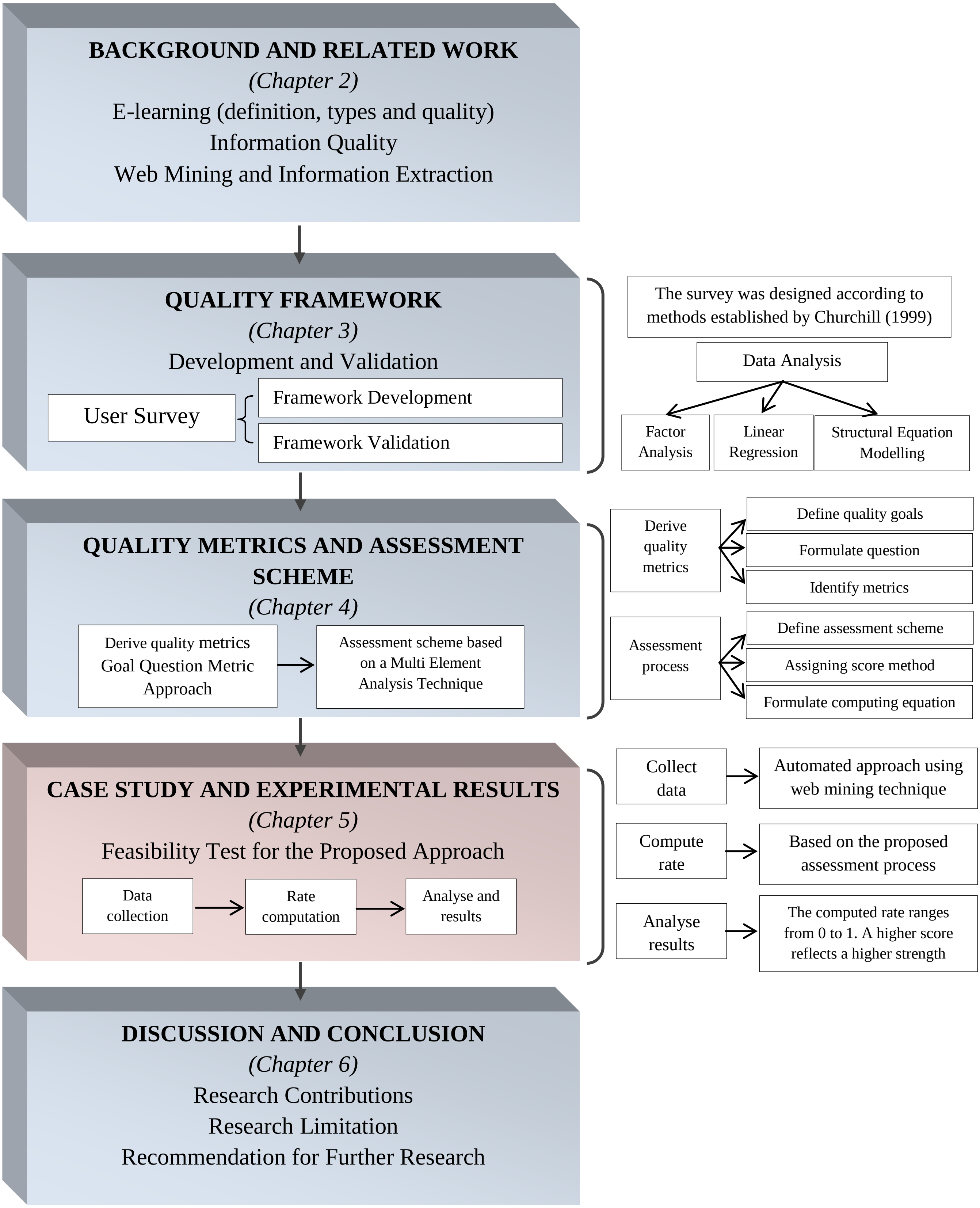


Figure 24. Thesis structure / Chapter 5

5.1 Introduction

This chapter focuses on the case study and the experimental results. It details the procedure of the technical experiment as a feasibility test for the proposed quality framework and the defined measurement approach. This chapter also includes data analysis to determine quality scores and feedback about the IQ, within a given e-learning system. Moreover, it shows how the results can be interpreted, in order to provide suggestions for future improvements and recommendations.

This chapter is organised as follows: the next section provides an overview of the experimental procedure. Details of the experiment will be discussed within the following sections, including the three main phases of the case study, namely, users' satisfaction survey, applying the proposed framework, which we proposed in Chapter3, and the automation of the measurement process. Then, we will compare results from the three phases. Finally, it will conclude with a detailed discussion of these results.

5.2 Experimental Plan and Procedure

This section summarises the main steps involved in our experimental plan and procedure. As illustrated in Figure 25, we started by selecting web pages to be evaluated. Then we followed a three-phase plan for data collection and quality score calculation.

Within the first phase, we conducted a user satisfaction survey to determine user evaluations of quality levels for information published on the nominated web pages; the results will be compared later with the results obtained from the next two phases.

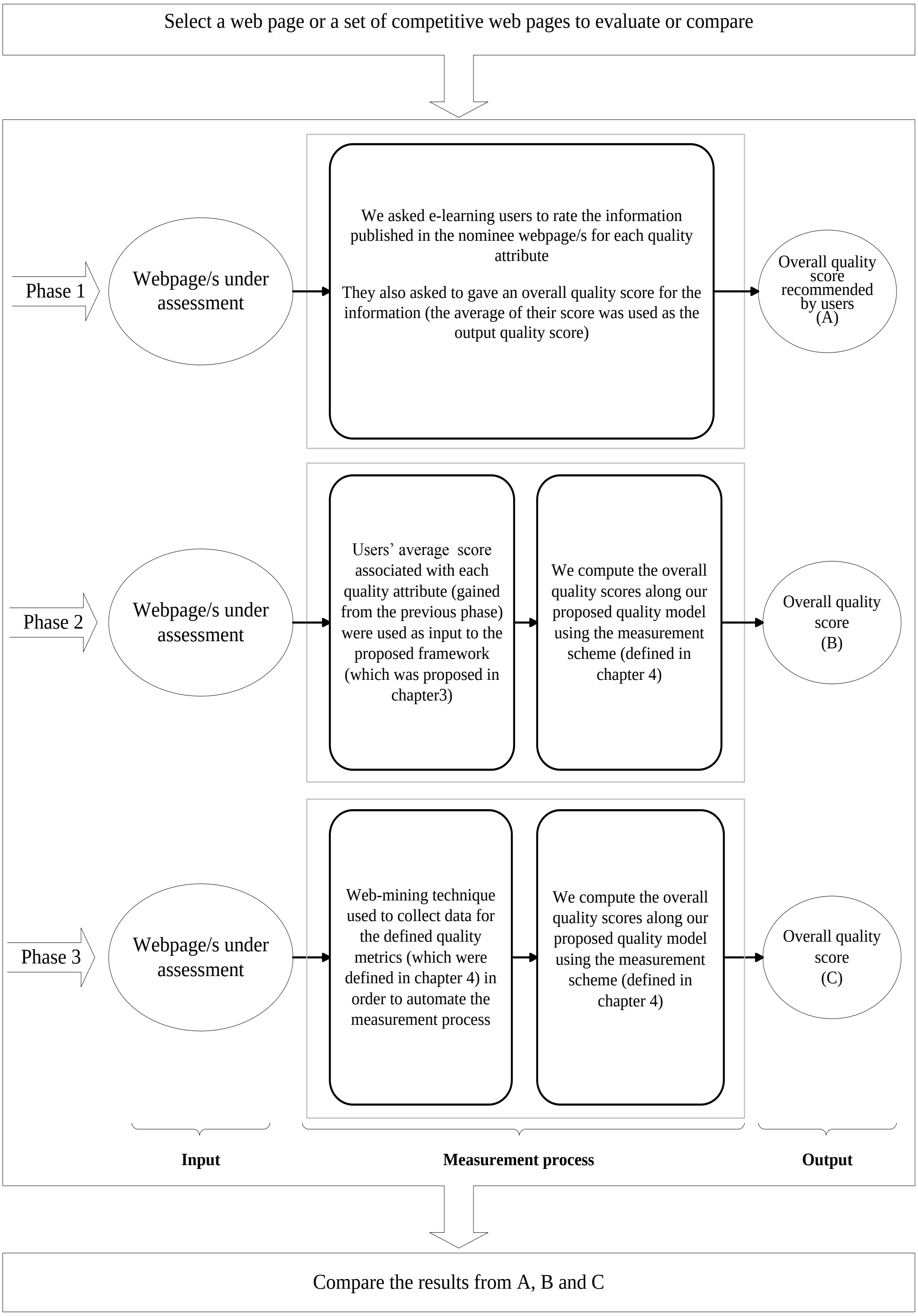


Figure 25. Case study procedure

Through this survey, we asked a group of e-learning users to rate the information for each quality attribute; they also were asked to give an overall quality score for the information. The average of their score was used as the first output for overall quality score (A).In the second phase, average user scores associated with each quality attribute (gained from the previous phase) were used as input for our framework. While in the third phase a web mining technique was used to collect data for the quality metrics (defined in Chapter 4) associated with each quality attribute, in order to automate the measurement process. In both phases, the overall quality scores were computed with our proposed quality framework using the measurement scheme, which was defined in Chapter 4, to reach the second and third outputs for overall quality scores (B and C). Then, we held a comparison between the results from previous phases.

5.3 Web Pages Selection and Keywords Identification

In order to test our proposed framework and measurement approach, we carried out a case study to evaluate, compare and rank information published in e-learning systems. Before starting our experiment we selected web pages to be examined and identified the necessary keywords.

5.3.1 Web Pages Selection

We selected two web pages from two mathematics e-learning systems which provided content in the same subject: set theory, particularly subsets.

The first web page (http://mathworld.wolfram.com/Subset.html) was taken from *MathWorld*™ website, a mathematical web resource provided by Wolfram Research. Its contents came out online in 1995, and targeted all educational levels [121].

The second web page (http://planetmath.org/encyclopedia/Subset.html), was taken from *PlanetMath* website, a mathematical web repository with a pedagogical aim. Its contents came out online in 2000 and the site was built by the users, representing a unique sort of hybrid of *MathWorld* and *Wikipedia* [122].

However, taking into account the fact that most websites evolve dynamically and frequently, and to make sure that the data collection process was not affected by any unexpected change of web page contents, we followed-up the content continuously. During the period of data collection (which began on 20 April, and finished on 30 May, 2010), we did not observe any major changes in content on these web pages that could have affected the evaluation process.

5.3.2 Keywords Definition

To identify the necessary keywords, which would be used within the quality metrics in the evaluation process, we collected five keywords associated with the web pages' subjects from two books on the same topic [123, 124]. The keywords are: subset, set, proper subset, empty set and power set.

Three mathematics lecturers were then consulted, they suggested adding another keyword: superset. So, we have six keywords in total for the measurement procedure namely; subset, set, proper subset, empty set, power set and superset.

5.4 User Satisfaction Survey

This section focuses on the first phase of our experiment where we conducted a user satisfaction survey to determine user evaluation of quality levels for information published on the selected two web pages. In the following subsections we will discuss survey design, participants' profiles, data collection methods and achieved results.

5.4.1 Survey

As mentioned before, in the first phase of the case study we conducted a user satisfaction survey, intended to measure the level of user satisfaction with IQ within two web pages. The survey consisted of three main sections, including a series of fixed-response statements and questions.

The first section obtained a concise profile of the respondent. While in the second section, we asked the respondents to rate the information published on the first web page (*mathworld.wolfram*), along the 14 quality attributes which were contained in the proposed framework, on a 10-point Likert scale (1 to 10), where lower scores indicated a lower quality level. Moreover, this section contained a general core question in which respondents gave an overall quality score for the published information. In the third section, we repeated the same questions, but they were to be answered with regard to the information published on the second web page (*PlanetMath*).

The survey was pre-tested by five respondents who were involved with computer science and mathematics research and use e-learning resources continuously. These respondents were not part of the final sample. Only minor changes were suggested.

In order to reduce costs and cut transfer times, the survey was distributed to the respondents via e-mail. The collected data was then transferred into an SPSS file in preparation for analysis[7].

[7] The survey is illustrated in Appendix IV.

5.4.2 Overview of Participants

The survey was carried out on a sample from a population of final-year bachelor degree students from Saudi university. These students were studying a mathematics course and dealt with e-learning systems on a regular basis. In May 2010, we collected sample data of 27 students, 12 of the responses were from men, while 15 of the respondents were female. Their ages ranged from 21 to 23 years of age.

5.4.3 Data Collection and Results

After analysing the data, it appears that users gave a higher quality score for the content on the second web page. While users awarded the overall quality of the first page (*mathworld.wolfram*) an average score of 0.8111, the average quality score for the second web page (PalnetMath) was 0.9111.

In more detail, Table 10 lists users' average scores associated with each quality attribute.

Quality attribute	Quality scores	
	mathworld.wolfram	***Planetmath***
Conciseness	0.8222	0.9222
Verifiability	0.6889	0.9222
Representational consistency	0.8333	0.9444
Understandability	0.7667	0.9111
Amount of information	0.8000	0.8556
Reputation	0.6556	0.7444
Completeness	0.6111	0.8667
Availability	0.7333	0.9000
Relevancy	0.8556	0.8667
Accessibility	0.9111	0.8889
Response time	0.9333	0.9000
Objectivity	0.8667	0.8667
Accuracy	0.8222	0.9222
Believability	0.8222	0.9333

Table 10. Users' average scores for each quality attribute

From this table we can see that users gave the second web page higher quality scores for most of the attributes, although the first web page had higher scores in accessibility and response time quality attributes; meanwhile, they had the same quality score for objectivity. These results will be used as the input for the second phase which will be discussed in the following section. Moreover, we will compare it later with the results from the next two phases.

5.5 Application of the Proposed Information Quality Framework

In this phase, we computed the overall quality scores using our proposed quality framework, using the measurement scheme and equation which were defined in Chapter 4. We used users' average scores, as associated with each quality attribute from the previous section, as input to populate the model.

Using the assigned relative importance weight for each attribute within the main quality factors, and a relative importance weight for each factor within the overall quality within the proposed framework, and depending on the defined measurement scheme, we calculated the overall quality score for each web page. The second web page, once again, recorded a higher quality score of 0.8881, while the first web page came second with an overall quality score of 0.8136.

5.6 Automation of Data Collection and the Measurement Process

For the third phase, an automated approach was applied, using a web mining technique, to collect the necessary data to assign values for the defined metrics within each quality attribute. Then, the quality score for each attribute was calculated as the average of the values of the related metrics. Finally, and similarly to what was done in the previous phase, the overall quality scores were computed along our proposed quality framework using our proposed measurement scheme.

In this section, a brief overview will first be given of the general plan. Followed by, discussions about data collection methods used for each quality attribute. Finally, this section will conclude with details of the results achieved at the end of the measurement process.

5.6.1 Overview of the General Plan

In order to automate the measurement process, Java programming techniques were used to mine the selected web pages, process the provided information and then extract the required data for the defined metrics.

To build the application[8], we used Java 2 Standard Edition (J2SE) Software Development Kit version 1.6 update 13, and Java API for XML Processing (JAXP) version 1.4.3. Also, we used Hyper Text Transfer Protocol. (HTTP) as the standards protocols to connect to web data sources. Within the applications, we adopted two packages "*myProject.parsing*" and "*myProject.query*", which was introduced by Tony Loton[9] in his book "*Web Content Mining*" [125], to parse the web pages and to pick up the desired HTML or XML elements. In addition, we built a new package "*myProject.quality*" to mine the results from the previous packages and extract the required data quality metrics associated with each quality attribute. Figure 26 shows a Unified Modelling Language (UML) outline of the Java applications arranged into Java packages.

[8] All codes are attached in a separate CD.
[9]With the kind permission of the author.

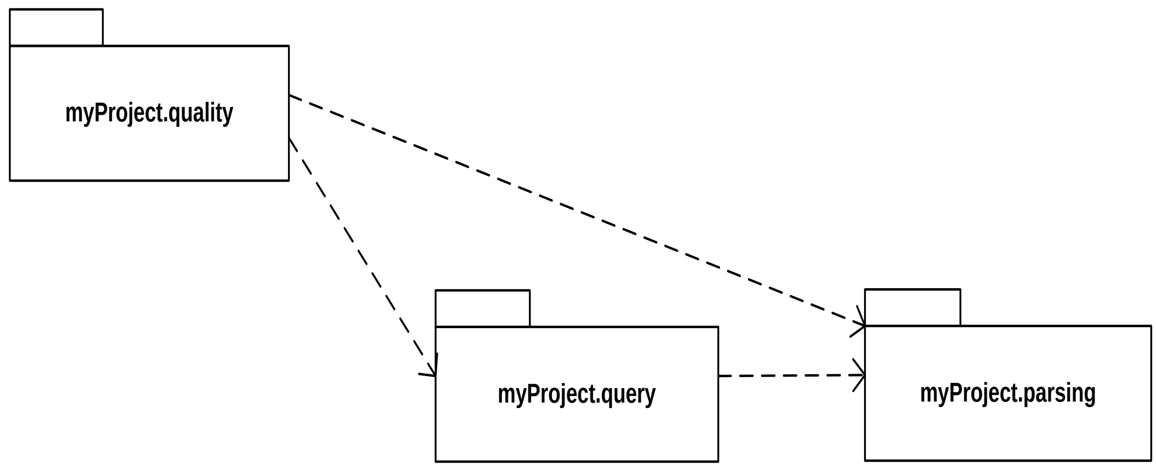


Figure 26. Utilised Java packages

Before digging deeply into the details regarding data collection methods for each quality attribute, first this section will give a general overview of each package and the corresponding classes:

5.6.1.1 The Parsing Package

As mentioned earlier, this package was first introduced by Tony Loton [125], and it was used as a core parsing technique upon which the two other packages depended.

Within this package there were three classes and an interface, as shown in Figure 27. The starting point is *WebParserWrapper* class, this class was used to submit the HTTP request and then examine the given content, HTML or XML, in order to decide upon the most suitable parser. Depending on the kind of content found, this class invokes the correct parser, *HTMLParserWrapper* or *XMLParserWrapper*. Both parsers implement a generic *WebParser* interface, so they can be used interchangeably.

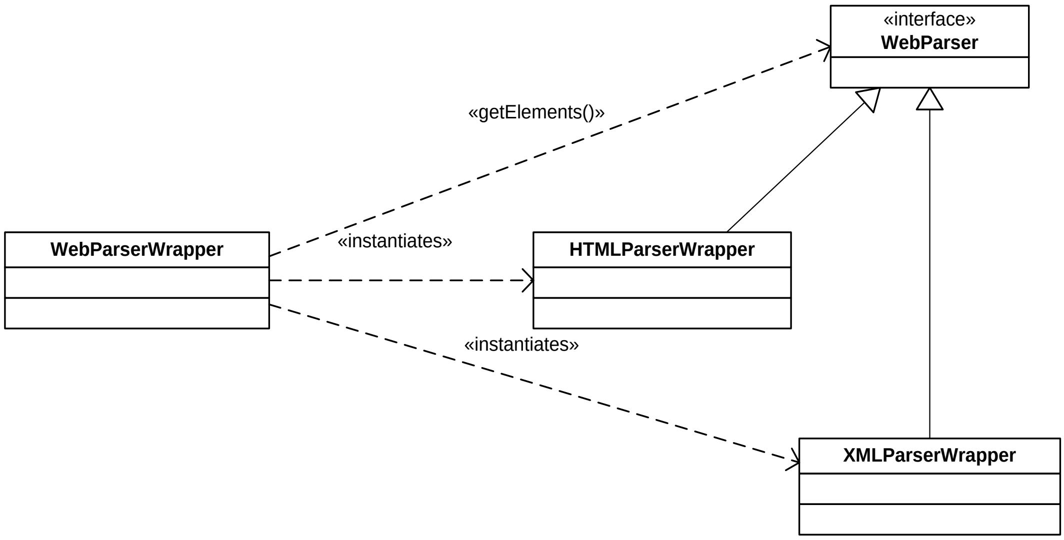


Figure 27. UML class diagram for package *myProject.parsing* Source [125 p 34]

5.6.1.2 The Query Package

This package was also adopted from the same book as the previous package [125]. It was used to apply filters and structure queries on the content which was parsed using the previous package. It enabled the picking out of particular HTML or XML elements, according to their content or positions within the mark-up structure.

This package contains five main classes and also utilised the *WebParserWrapper* class from the first package; the UML class diagram for this package is shown in Figure 28. In order to enable the filtering, first the author generated an *Operator* class to carry out basic numeric and lexical comparison operators for data items. This class will be used by a *Filter* class to apply the required filters on the parsed content. In order to view the filtered results there is a *FilterViewer* graphical class. The filtering concept is extended further by using *QueryEngine* and *SqlGui* classes. However, practically, in our applications we only took advantage of the *Operator* from this package, as will be explained later in this chapter.

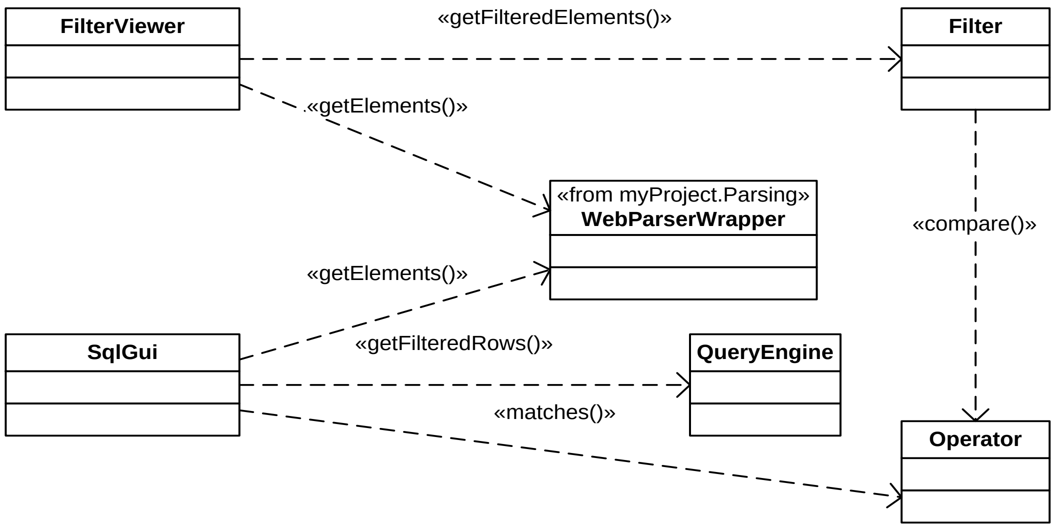


Figure 28. UML class diagram for package *myProject.query* Source [125 p. 68]

5.6.1.3 The Quality Package

This package was built to be the basis for extracting the required information from the parsed content. This package contained eleven main classes and also utilised the *WebParserWrapper* and the *Operator* classes from the first and the second packages, respectively.

The UML class diagram for this package is shown in Figure 29. It includes classes to collect the required information to be used for assigning metrics' scores and calculating quality values for ten quality attributes, namely; accessibility, accuracy, amount of information, believability, completeness, conciseness, consist, objectivity, understandability and verifiability. As a result of space restrictions and to make the diagram in Figure 29 more visible and readable, we used the term *"Qualityattribute(i)"* to express these ten classes. These defined classes invoke the *getElement()* method from the *WebParserWrapper* class, in order to submit the HTTP request for a given web page and parse the discovered content. Then, the *matches()* method from the *Operator* class was called upon to carry out a wildcard string comparison between the parsed content and a pre-defined template. Moreover, an *OverallQuality* class was defined which instantiated the other ten classes and computed the overall quality score along the proposed quality framework using the assigned relative importance values.

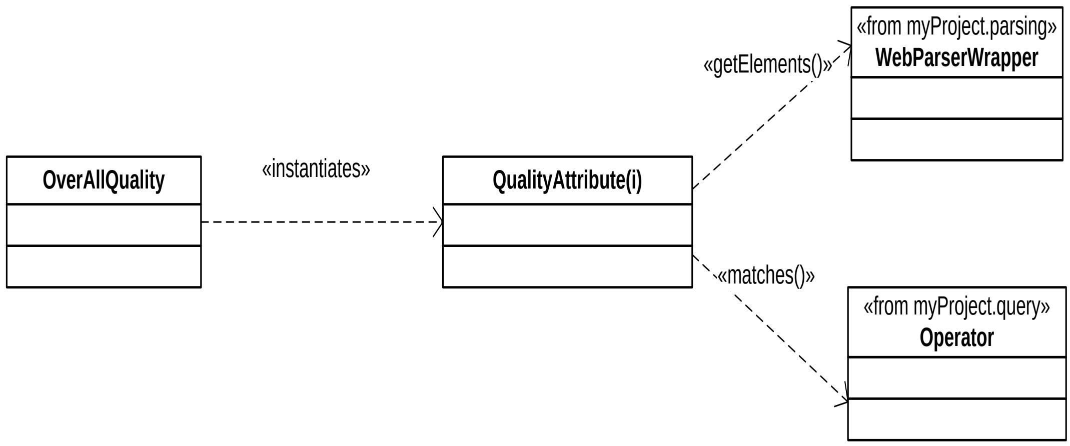


Figure 29. UML class diagram for package *myProject.quality*

Within the next two sections more details will be provided regarding data collection methods for each quality attribute and then the accomplished results will be presented.

5.6.2 Data Collection Methods for Each Quality Attribute

This section will discuss data collection methods for each quality attribute. As mentioned previously, the quality package contains classes to calculate quality scores for ten quality attributes. For the remaining four attributes (reputation, availability, relevancy and response time) online tools were used to calculate the values for associated metrics. The following two subsections contain more detailed explanations of classes and tools used for each individual quality attribute.

5.6.2.1 Quality Attributes for which Quality Scores were Calculated Using Java Classes

In the following, a detailed explanation is given, in the form of tables, for the classes which were used to assign the quality scores for ten quality attributes[10].

– ***Conciseness***

Class	Methods	Parameters	Description
Conciseness	***Vector*** *getFeatureString()*	***Vector*** inputwords ***string*** strings	This method steps, through a set of input strings, invokes the *match()* method from *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Vector *getFound()*	***Vector*** Input **Vector** templates	This method steps, through a set of input sentences, calls upon *getFound()* to identify links and text and then invoke the *match()* method from the *Operator* class to carry out wildcard comparisons between these sentences and a set of templates. It returns a vector of matching sentences.
	String *tagHerf()*	***String*** input	This method picks out each sentence from input sentences and identifies links and text within these sentences and tags them and returns a vector of tagged sentences.
	Double *getCons()*	***String*** pageURL ***String[]*** keywords	This method takes a web page's URL and an array of keywords as parameters, then it invokes *WebParserWrapper* class to parse the content of this page. It also calls upon the other three methods to identify the hyperlinked keywords and calculate and return the conciseness score.

Output:
The final output from this class will be the conciseness score as a double value range between 0 and 1. A higher value reflects a better quality score.

Table 11. Summary of Conciseness class which was used to assign the quality score for Conciseness quality attribute

[10] The full documentation for the quality package could be found in Appendix V.

– *Verifiability*

Class	Methods	Parameters	Description
Verifiability	***Vector*** *getFeatureString()*	***Vector*** inputwords **string** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Double *getVer()*	***String*** pageURL	This method takes a web page's URL as a parameter, then it invokes *WebParserWrapper* class to parse the content of this page. It also calls upon *getFeatureString()* to check if there are any means of verifying information and returning the verifiability score.

Output:
The final output from this class will be the verifiability score, which will be either 1 if the web page provides links to verify the information or 0 if not.

Table 12. Summary of Verifiability class which was used to assign the quality score for Verifiability quality attribute

– *Understandability*

Class	Methods	Parameters	Description
Understandability	***Vector*** *getFeatureString()*	***Vector*** inputwords ***string*** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings
	Double *getUnder*	***String*** pageURL ***String[]*** keywords	This method takes a web page's URL as a parameter, then it invokes the *WebParserWrapper* class to parse the content of this page. It also calls upon *getFeatureString()* to check if there are any means of explanation components and returns the understandability score.

Output:
The final output from this class will be the understandability score, which will be either 1 if the web page provides an explanation component or 0 if not.

Table 13. Summary of Understandability class which was used to assign the quality score for Understandability quality attribute

- ***Representational consistency***

Class	Methods	Parameters	Description
Consist	***Boolean*** *checkConsistency()*	***Vector*** input1 ***Vector*** input2	This method steps inside two sets of metadata from two web pages, picks the style sheets, links and compares them. It returns true if the style sheets match and false if not.
	Double *getConsis()*	***String*** pageURL	This method takes a web page's URL as a parameter and identifies the domain's URL. Then it invokes *WebParserWrapper* class to parse the content of the two URLs. It also calls upon *checkConsistency()* to compare the style sheets and return the representational consistency score.
Output: The final output from this class will be the representational consistency score, which will be either 1 if the web page has the same style sheet as the homepage or 0 if not.			

Table 14. Summary of Consist class which was used to assign the quality score for Representational consistency quality attribute

- ***Accuracy***

Class	Methods	Parameters	Description
Accuracy	***Vector*** *getFeatureString()*	***Vector*** inputwords ***string*** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Double *getAcc()*	***String*** pageURL	This method takes a web page's URL as a parameter, then it invokes *WebParserWrapper* class to parse the content of the URL. It calls *getFeatureString()* to check whether the content contains any bibliography, whether there is any indication for the last update and whether it provides any additional resources. It then returns the accuracy score.
Output: The final output from this class will be the accuracy score as a double value range between 0 and 1. A higher value reflects a better quality score.			

Table 15. Summary of Accuracy class which was used to assign the quality score for Accuracy quality attribute

– ***Amount of information***

Class	Methods	Parameters	Description
AmountofInformation	***Vector*** *getFeatureString()*	***Vector*** inputwords: ***string*** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Double *getAOI()*	***String*** pageURL ***String[]*** Nkeywords ***String[]*** Pkeywords	This method takes a web page's URL and two arrays of needed keywords and provides keywords as a parameter*, then it invokes *WebParserWrapper* class to parse the content of this page. It also calls upon *getFeatureString()* to check how many necessary keywords were found in the content, then returns an information score.

Output:

The final output from this class will be the amount of information score as a double value range between 0 and 1. A higher value reflects a better quality score. It also raises a flag to indicate if the information is too little or too much.

*The provided keywords are identified from the web page using Google Keywords Destiny tools (http://googlerankings.com/ultimate_seo_tool.php).

Table 16. Summary of AmountOfInformation class which was used to assign the quality score for Amount of information quality attribute

– ***Completeness***

Class	Methods	Parameters	Description
Completeness	***Vector*** *getFeatureString()*	***Vector*** inputwords **string** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Double *getCom()*	***String*** pageurl ***String[]*** keywords	This method take a web page's URL and an array of keywords as parameter, then it invokes the *WebParserWrapper* class to parse the content of this page. It also calls upon *getFeatureString()* to check how many keywords were found in the content, then returns the completeness score.

Output:

The final output from this class will be the completeness score as a double value range between 0 and 1. A higher value reflects a better quality score.

Table 17. Summary of Completeness class which was used to assign the quality score for Completeness quality attribute

– ***Accessibility***

Class	Methods	Parameters	Description
Accessibility	***Vector*** *getFeatureLinks()*	***Vector*** links	This method steps through a set of hyperlinks, it calls on *isLive()* to check broken links and return a vector of all working links.
	Boolean *isLive()*	***String*** link	This method uses a hyperlink as a parameter and returns it as false if the link is broken and true if not.
	Vector getLinks()	***Vector*** input	This method picks out all hyperlinks from a vector of input sentences and returns a vector of these links.
	Double *getAccess()*	***String*** pageURL	This method takes a web page's URL as a parameter, then it invokes *WebParserWrapper* class to parse the content of this page. It also calls upon *getFeatureLinks() and getLinks()* to check how many links are broken, then returns the accessibility score

Output:
The final output from this class will be the accessibility score as a double value range between 0 and 1. A higher value reflects a better quality score.

Table 18. Summary of Accessibility class which was used to assign the quality score for Accessibility quality attribute

– ***Believability***

Class	Methods	Parameters	Description
Believability	***Vector*** *getFeatureString()*	***Vector*** inputwords ***string*** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Double *getBel()*	***String*** pageURL	This method takes a web page's URL as a parameter, then it invokes *WebParserWrapper* class to parse the content of the URL. It calls *getFeatureString()* to check whether the content contains any information about the authors and publisher body and if it provides any contact information. It then returns the believability score.

Output:
The final output from this class will be the believability score as a double value range between 0 and 1. A higher value reflects a better quality score.

Table 19. Summary of Believability class which was used to assign the quality score for Believability quality attribute

– ***Objectivity***

Class	Methods	Parameters	Description
Objectivity	***Vector*** *getFeatureString()*	***Vector*** inputwords ***string*** strings	This method step, through a set of input strings, invokes the *match()* method from the *Operator* class to compare these strings against a string. It returns a vector of matching strings.
	Double *getLinks()*	***Vector*** input	This method extracts all hyperlinks from a set of sentences, checks the domain extensions and counts the links with unbiased links. Then it returns the number of these links.
	Double *getObj()*	***String*** pageURL	This method takes a web page's URL as a parameter, it checks the domain extension, then it invokes *WebParserWrapper* class to parse the content of the URL. It calls on *getFeatureString()*to check if there is any information about the publisher body. It also calls on *getLinks()* to find the number of unbiased links within the page then returns the objectivity score.

Output:

The final output from this class will be the objectivity score as a double value range between 0 and 1. A higher value reflects a better quality score.

Table 20. Summary of Objectivity class which was used to assign the quality score for Objectivity quality attribute

5.6.2.2 Quality Attributes for which Quality Scores were Calculated Using Online Tools

This subsection provides explanations for the tools, which were used to calculate the quality scores for the four remaining attributes. As in the previous subsection, the explanation is arranged in the form of tables.

However, although these well-known tools were widely used in order to save time, it is important to consider building our own calculation methods as a future work.

– ***Availability and response time***

Tool	Description	Usage
Mon.itor.us online web monitoring tool Available from the following link: http://www.mon.itor.us	http://mon.itor.us was launched as a website monitoring service in March 2006, it provides, for its users, a 24 x 7 network and website monitoring service. It also could be used as a research tool for web traffic correlations with uptime and performance.	Checks the availability and the performance of specific websites/web pages and gives daily e-mail reports.
Output: The output will be availability and response time scores ranged between 0 and 1. A higher value reflects a better quality score.		

Table 21. Explanations for the tool, which was used to calculate the quality scores for Availability and Response time quality attributes

– ***Relevancy***

Tool	Description	Usage
Free Monitor for Google Version 2.5.28.75 Available to download from: http://www.cleverstat.com/en/google-monitor-query.htm	A free search engine position software designed to send a query to Google and show the position of a website by specific target keywords.	Search engine position software which finds the position of a specific website/page in Google Top for specific keyword/s
Output: We will consider that the score for each keyword search is 1 if we find the web page within the first three pages in Google search results, and 0 if not*. The overall relevancy score is the average of the score for all the keywords.		
* We decided to look within the first three pages, depending on a research finding which found that 90% of search engine users click on results within the first three pages of search results.[126]		

Table 22. Explanations for the tool, which was used to calculate the quality scores for Relevancy quality attribute

– ***Reputation***

Tool	Description	Usage
Page Rank Checker Available from the following link: http://www.prchecker.info/check_page_rank.php	A trademark of Google Inc. which uses Google PageRank™ algorithm to check the ranks of all website pages	A free tool to check a Google™ page ranking of any website page

Output:
The output will be a reputation score ranging between 0 and 1. A higher value reflects a better quality score.

Table 23. Explanations for the tool, which was used to calculate the quality scores for Reputation quality attribute

5.6.3 Results

After applying the automated approach, as explained earlier in this section, we obtained an overall quality score for the information published on each of the selected web pages, as well as a score for each quality attribute within our proposed framework. For the third time, the overall quality score associated with *PlanetMath* (0.8559) was higher than the score reached for *MathWorld* (0.8273).

In contrast to the results from the first phase, when looking at scores associated with each quality attribute listed in Table 24, we can see that *MathWorld* recorded higher scores for eight quality attributes. Nevertheless, *PlanetMath* had a higher score in the objectivity quality attribute, while they recorded the same quality scores for accessibility, representational consistency, verifiability, response time and understandability attributes.

In the next section, we will provide a more detailed comparison of the results from the three phases of the case study.

Quality attribute	**Quality score**	
	mathworld.wolfram.com	***Planetmath.org***
Conciseness	1.0	0.6
Verifiability	1.0	1.0
Representational consistency	1.0	1.0
Understandability	1.0	1.0
Amount of information	0.8333	0.6667
Reputation	0.8	0.6
Completeness	1.0	0.8333
Availability	1.0	0.9931
Relevancy	0.8333	0.6667
Accessibility	1.0	1.0
Response time	1.0	1.0
Objectivity	0.3333	0.9974
Accuracy	1.0	0.6667
Believability	1.0	0.6667

Table 24. Recorded quality score for each quality attribute from third phase

5.7 Results Comparison

This section will provide a comparison of the results achieved from the three phases of the case study, however, more detailed explanation and interpretation of these results will follow in the discussion and conclusions section.

Table 25 shows the overall quality scores achieved from the three phases of the case study, which were explained in this chapter. It can quite clearly be seen that *PlanetMath* recorded the highest scores in the three phases.

Web page	**Overall quality score**		
	First phase (users)	***Second phase (model)***	***Third phase (automatic)***
mathworld.wolfram.com	0.8111	0.8136	0.8273
Planetmath.org	0.9111	0.8881	0.8559

Table 25. Overall quality scores recorded in the three phases of the case study

In order to investigate differences between the overall quality scores reached from the three phases, we conducted a paired t-test, using SPSS, to determine statistical significances within these differences [101].

Table 26 presents the output for the t-test conducted to examine differences between the recorded overall quality scores from the first and second phases, and first and third phases respectively. The table contains the mean for quality scores, mean differences and the two-tailed p values for difference significance. Based on Field and Babbie, any value less than 0.05 for the two-tailed p values indicates a significant statistical difference between the two scores [102, 103].

Derived from the p values presented in the table, we can state that the differences between the overall quality scores, as recorded within the different phases in our case study, are not significant enough to be taken into account.

	Mean	t	
		P=Sig. (2-tailed)	**Mean Difference**
Users ***Model***	.861100 .850850	.569*	.01025
Users ***Automatic***	.861100 .841600	.682*	.0195

***** *p>0.05 No significant difference.***

Table 26. T-test results for overall quality scores

In terms of the differences between the recorded scores associated with each quality attribute from the first and third phases, Table 27 summarises these scores for each examined web page.

Quality attribute	Quality score			
	mathworld.wolfram.com		***Planetmath.org***	
	Users	**Automatic**	**Users**	**Automatic**
Conciseness	0.8222	1.0	0.9222	0.6
Verifiability	0.6889	1.0	0.9222	1.0
Representational consistency	0.8333	1.0	0.9444	1.0
Understandability	0.7667	1.0	0.9111	1.0
Amount of information	0.8000	0.8333	0.8556	0.6667
Reputation	0.6556	0.8	0.7444	0.6
Completeness	0.6111	1.0	0.8667	0.8333
Availability	0.7333	1.0	0.9000	0.9931
Relevancy	0.8556	0.8333	0.8667	0.6667
Accessibility	0.9111	1.0	0.8889	1.0
Response time	0.9333	1.0	0.9000	1.0
Objectivity	0.8667	0.3333	0.8667	0.9974
Accuracy	0.8222	1.0	0.9222	0.6667
Believability	0.8222	1.0	0.9333	0.6667

Table 27. Scores associated with each quality attribute from the first and third phases of the case study

Although to some extent the scores seem to be quite similar, once again, we conducted a paired t-test to determine whether there were any statistically significant differences between the results achieved from the two phases. The results derived from this test are listed in Table 28.

Quality attributes	Mean		t	
	Users'	***Automati***	***P=Sig. (2-tailed)***	***Mean Difference***
Conciseness	.872200	.800000	.821*	0.0722
Verifiability	.805550	1.00000	.344*	-0.19445
Representational consistency	.888850	1.000000	.295*	-0.11115
Understandability	.838900	1.000000	.268*	-0.1611
Amount of information	.827800	.750000	.611*	0.0778
Reputation	.700000	.700000	1.000*	0
Completeness	.738900	.916650	.555*	-0.17775
Availability	.816650	.996550	.286*	-0.1799
Relevancy	.861150	.750000	.429*	0.11115
Accessibility	.900000	1.000000	.070*	-0.1
Response time	.916650	1.000000	.126*	-0.08335
Objectivity	.866700	.665350	.653*	0.20135
Accuracy	.872200	.833350	.887*	0.03885
Believability	.877750	.833350	.874*	0.0444

* $p>0.05$ No significant difference.

Table 28. T-test results for scores associated with each quality attribute

5.8 Discussions and Conclusions

In order to evaluate, compare and rank IQ in e-learning systems, we specified 14 quality attributes within our proposed framework. Then, we identified associated metrics and defined the measurement scheme. In this chapter, a case study was carried out as a feasibility test for the proposed quality framework and the defined measurement approach.

To complete this experiment it was important to go through three basic phases as shown in Figure 25. First a users' satisfaction survey was conducted, where a group of e-learning users were asked to rate the information within two selected web pages, for each quality attribute, and they also were asked to give a score for the overall quality of the information on each web page.

Following this, in the second phase, the proposed quality framework was populated with the average scores for each quality attribute, which were recorded in the first phase. While in the third phase, the model was populated with quality scores calculated automatically using a web mining technique, to collect the required data to be used within the identified quality metrics. In order to determine the overall quality score in both phases, the defined measurement scheme was applied. Finally, a comparison was made between the results from the three phases.

Although results and findings from the case study are limited by the relatively small sample size and the fact that the study only considered two web pages, it does provide an initial exploratory idea about the feasibility of the proposed quality measurement approach and this can be used to draw conclusions. In fact, the comparison results indicate that the proposed measurement model and scheme could give a good indication about IQ level in e-learning systems.

Looking to Table 25, from the previous section, it can clearly be seen that the overall information quality scores gained from the second phase were closer to the scores given by users than the results from the third phase. Moreover, when considering the t-test results in Table 26, although it was not possible to notice any significant differences when comparing quality scores from the second and third phases, with those recorded in the first phase, it can be perceived that the difference between quality scores given by the users and calculated using the proposed quality model was less than between those calculated automatically. Consequently, it can be stated that using the proposed quality framework could give a good indication of the information quality in an e-learning context, which could be very useful for e-learning system developers and users.

Regarding the automated approach, Table 27 showed detailed results for the differences between the recorded scores associated with each quality attribute from the first and third phases. Although the scores look quite similar, there are some clear differences in certain quality attributes associated with each web page, such as completeness and objectivity for the *MathWorld* web page and conciseness, accuracy and believability for the *PlanetMath* web page. It is true that the t-test results in Table 28 showed that these differences are not statistically significant; nevertheless, it is possible to see from the mean differences that there is space to improve the automated approach for a number of quality attributes, such as verifiability, understandability, completeness, availability and objectivity. We believe that these improvements should be made mainly in the definition of the quality metrics. This is because the only difference between the second and the third phases was in the method of calculating the input scores, otherwise the same measurement scheme was followed.

To sum up, it is possible to conclude that the results of the conducted case study reveal that the proposed model for information quality in an e-learning context could be used as an overall IQ indicator. Moreover, although the automated approach could also give good results, it could be improved to mimic the users' view points.

Indeed, there should be more experiments to determine and highlight which parts of the defined metrics are responsible for the existence of differences

Chapter 6: Conclusions and Recommendations

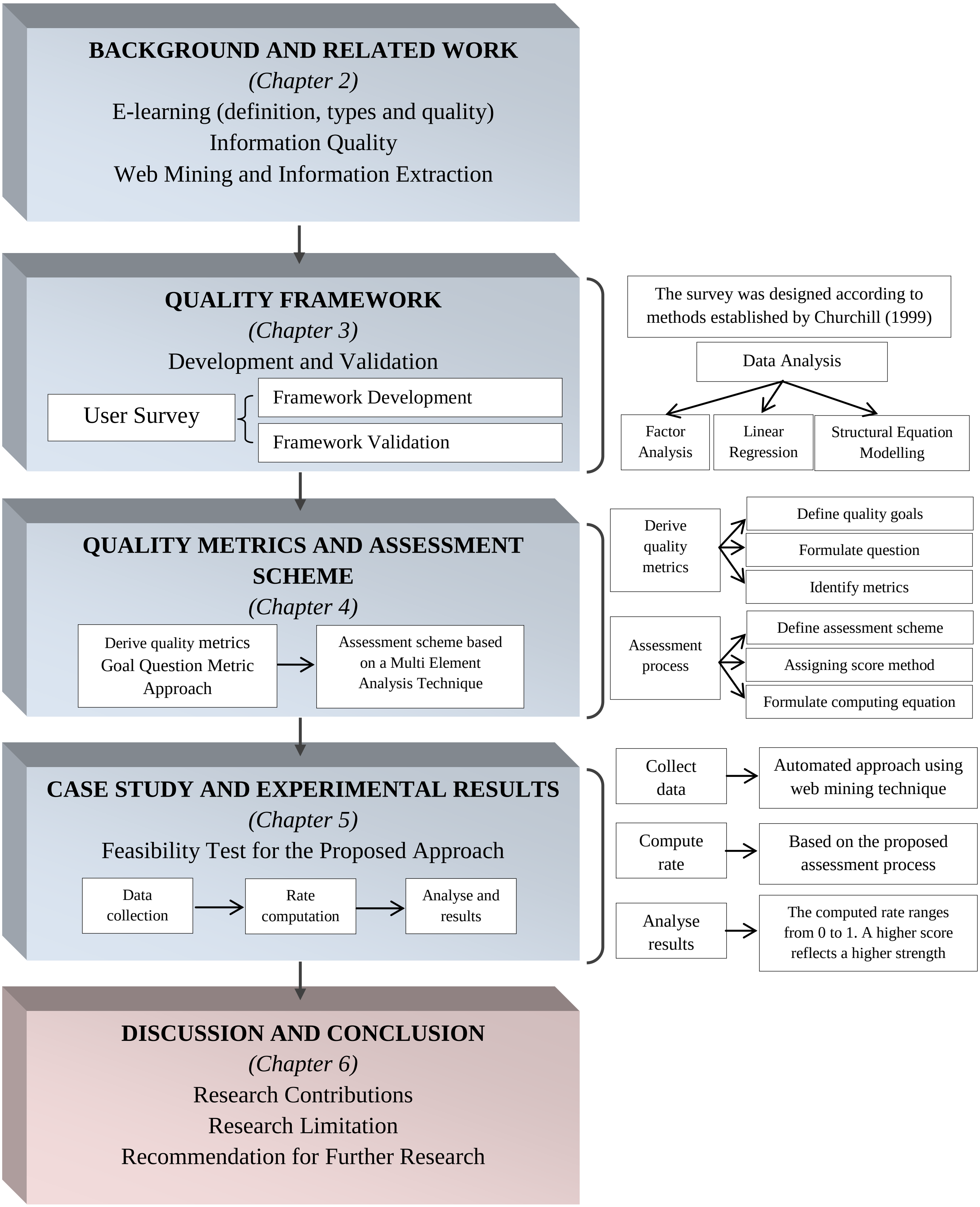


Figure 30. Thesis structure / Chapter 6

The aim of our study was to set quality standards for the content provided by e-learning systems, by defining the main dimensions of the information quality. Besides that, it focused on identifying a set of metrics to quantify the quality of the information, in order to facilitate the evaluation, comparison and analysis of information quality. Moreover, because human judgment is fallible, this research aimed to examine the feasibility of integrating web mining techniques as a means of gathering the necessary information to conduct the evaluation measurement. The feedback from analysing the information quality scores for a given e-learning system can provide suggestions and recommendations for future improvements.

The study started by reviewing the earlier proposed frameworks, from the relevant literature; the identified quality attributes and characteristics were used together with a user survey in order to develop and validate a framework, which represented factors and attributes that impact upon information quality in e-learning systems. The main structure of the proposed framework was based on Wang and Strong's DQ framework [35].

After that, a goal question metric approach was used [110] as a mechanism for defining and interpreting operational and measurable quality metrics. Then, a multi element analysis technique was used to define the assessment scheme [117]. Finally, a case study was carried out as a feasibility test for the proposed quality framework and the defined measurement approach.

The conclusions of this thesis will summarise research contributions, highlight limitations of the study and provide suggestions and recommendations for further research.

6.2 Research Contributions

This section revisits the main research questions and shows the achieved contributions regarding each question.

– ***Research contributions regarding the first research question***

"How can the key dimensions for information quality be identified, from the users' perspective, in order to build a quality framework to measure the quality of the content provided by e-learning systems?"

- This thesis contributes to the literature of IS, particularly in the field of IQ. It provides a comprehensive review for a number of the major historical developments of IQ frameworks (Chapter 2). Moreover, it provides a summary of the most common quality attributes between the examined frameworks, which could provide a good basis for IQ researches (Chapter 3). This review was published in [127]

- This thesis identified the key attributes for IQ from the users' perspective, within a proposed quality framework, to measure the quality of the content provided by e-learning systems; this consisted of 14 quality attributes grouped in three quality factors: intrinsic, contextual representation and accessibility, our work regarding the framework development was published in [128]. The framework has the flexibility to be adopted and used in different e-learning environments and with different users, as we used the relative importance as a parameter for the measurement process within the proposed framework, which could be updated if a quality attribute or factor appeared no longer to have the same significant value for the quality assessment. The framework also could be useful to shed some light on gender and cultural related differences in user perceptions of the relative importance of the main quality factors within the proposed framework [129]. Moreover, the thesis presented an empirical validation of the framework (Chapter 3). Validation methodology and results was published in [130]

– ***Research contributions regarding the second research question***

"How can a specified set of metrics be determined, to quantify the quality of information in e-learning systems, in such way that they will enable the evaluation, comparison and analysis of information quality in such systems?"

- Within the thesis quality metrics were developed for measuring IQ in an e-learning context. The GQM approach was used as a mechanism for defining and interpreting operational and measurable quality metrics (Chapter 4). Quality metrics definition was published in [131]

– ***Research contributions regarding the third research question***

"What is the most appropriate and applicable assessment scheme, used to compute the identified metrics, and which will ultimately reach an overall information quality assessment for the published materials?"

- We also defined an assessment scheme and process, based on a multi-element analysis technique. A detailed description of the measurement scheme was given, along with the scoring method and the equation used to compute quality scores (Chapter 4). All the details for the assessment scheme definition are included in [132]

– ***Research contributions regarding the fourth research question***

“How can web mining technologies be positively utilised, in order to automate the data collection and evaluation processes?"

- Within the thesis a case study was carried out as a feasibility test for the proposed quality framework and the defined measurement approach. Moreover, the case study aimed to observe the possibility of integrating web mining techniques as a means of gathering the required information to conduct the evaluation measurement. The results of the conducted case study reveals that the proposed framework could be used as an overall IQ indicator, while the automated approach, which could also give good results, could be improved to mimic the users’ view points, the details and the results from the case study are included in [132]

This research gives unique contributions because it proposed a solution to the problem resulting from an absence of consensus regarding evaluation standards and methods for measuring information quality in e-learning systems. Moreover, it suggested taking advantage of web mining techniques to automate the retrieval process of the information required for the quality measurement, in order to deliver certain and accurate quality scores, and also to reduce the level of effort and time spent.

In addition, this study opens up new directions for further research. However, before giving suggestions for future work, first a summary will be provided of the research limitations.

6.3 Research Limitations

Although the results can be considered promising and positive, and the framework and assessment scheme were revealed to be good predictors for IQ within e-learning systems, the research has some limitations which should be highlighted.

- The limitations of this study include the length of time for the study; a short time frame resulted in the most of the limitations.
- One result of time constraints was that a number of well-known online tools were used to calculate the quality scores for the four quality attributes during the case study, instead of building our own calculation methods.
- Once again, as a result of time limitations the case study was applied on only two single static web pages.
- There should be more experiments to determine and highlight which parts of the defined metrics are responsible for the existence of the differences.

6.4 Recommendations for Further Research

This thesis essentially covered two main areas of research: development and validation of an information quality framework for e-learning systems, and identification of a set of measurement metrics and defining a suitable measurement scheme. Moreover in this thesis we carry out a feasibility test for the proposed quality framework and the defined measurement approach by mean of a case study.

Further to the work reported in this thesis, it is suggested that there could be several advances for further research and development:

Possible further research includes updating the proposed framework continually by examining newly published literature to discover any new quality attributes. Furthermore, it is suggested that an extensive gender and cultural comparison study be undertaken in order to generalise the findings.

The feasibility test for the proposed quality framework and the defined measurement approach within this thesis is based on a single case study. Therefore, more experiments and case studies are suggested to determine and highlight which parts of the defined metrics are responsible for the existence of differences.

As mentioned earlier, as a result of time constraints we used a number of well-known online tools to calculate the quality scores for the four quality attributes during the case study. Although the utilised online tools gave good results, building our own calculation methods should be considered for future work.

Another suggested future work, would involve applying the defined measurement methods for a whole dynamic e-learning system and utilized the software agents' technology as a mean of information extraction methods in order to automate the retrieval process of the required information for the quality measurement. Using this technology could be beneficial for the assessment a dynamic e-learning system as it enables dynamic compatibility and allows interaction.

References

[1] R. C. Close, R. Humphreys, and B. W. Ruttenbur, *"E-learning & technology: Technology & the internet are changing the way we learn"* Sun Trust Equitable Securities, USA, 2000.

[2] G. Buyukozkan, D. Ruan, and O. Feyzioglu, "Evaluating e-learning web site quality in a fuzzy environment" *International Journal of Intelligent Systems*, vol. 22, pp. 567-586, May, 2007.

[3] C. M. Stracke, "Quality standards for quality development in e-learning: Adoption, implementation and adaptation." in proceedings of *World Conference on Educational Multimedia, Hypermedia and Telecommunication*, pp. 4158-4165, Chesapeake, VA, 2007

[4] M. P. Bozula and I. A. Morgun, 2008,Problems of distance courses materials expertise, *Scientific works of Vinnytsia National Technical University, Electronic scientific specialized edition*, [Online] vol. 4 Available:http://www.nbuv.gov.ua/e-journals/VNTU/2008-4/2008-4_en.htm.

[5] A. M. Aladwani and P. C. Palvia, "Developing and validating an instrument for measuring user-perceived web quality" *Information & Management*, vol. 39, pp. 467-476, May, 2002.

[6] J. Churchill, *Marketing research: Methodological foundation* (7th .ed), Hinsdale, IL: The Dryden Press, 1999.

[7] D. Harrington, *Confirmatory factor analysis*, (1st .ed), New York: Oxford University Press, 2009.

[8] B. Byrne, *Structural equation modeling with amos: Basic concepts, applications and programming* (1st .ed), Mahawa, New Jersey: Psychology Press, 2001.

[9] H. Koziolek, "Goal, question, metric" chapter in: *Dependability metrics: Advanced lectures*, I. Eusgeld, F.C. Freiling, and R. Reussner, Berlin: Springer-Verlag Berlin, 2008, pp. 39-42.

[10] K. M. Khan, "Assessing quality of web based systems" in proceedings of *International Conference on Computer Systems and Applications (IEEE/ACS)*, pp. 763-769, Doha, Qatar, March 31-April 04, 2008

[11] M. F. Paulsen, 2002,Online education systems: Discussion and definition of terms, *NKI Distance Education*, [Online] Available:http://www.nettskolen.com/forskning/Definition%20of%20Terms.pdf.

[12] A. Romiszowski, "How's the e-learning baby? Factors leading to success or failure of an educational technology innovation" *Educational Technology*, vol. 44, pp. 5-27, 2004.

[13] J. Gerhard and P. Mayr, "Competing in the e-learning environment--strategies for universities" in proceedings of *35th Annual Hawaii International Conference on System Sciences (HICSS'02)*, pp. 262-272, Big Island, Hawaii 7-10 January, 2002

[14] Y. Psaromiligkos and S. Retalis, "Re-evaluating the effectiveness of a web-based learning system: A comparative case study" *Journal of educational multimedia and hypermedia*, vol. 12, pp. 5-20, 2003.

[15] S. Naidu, 2006, *E-learning: A guidebook of principles, procedures and practices* [Online]. Available: http://www.cemca.org/e-learning_guidebook.pdf.

[16] Commission of the European Communities, 2001,The elearning action plan designing tomorrow's education, *Communication from the Commission to the Council and the European Parliament*, [Online] Available:http://www.europa.eu/eur-lex/en/com/cnc/2001/com2001_0172en01.pdf.

[17] B. Holmes and J. R. Gardner, *E-learning: Concepts and practice* (1st .ed), London: Sage Publications Ltd, 2006

[18] A. Rosen, *E-learning 2.0: Proven practices and emerging technologies to achieve real results* (1st .ed), New York: AMACOM (American Management Association), 2009.

[19] D. Spencer and S. R. Hiltz, "Studies of aln: An empirical assessment" in proceedings of *34th Hawaii International Conference on System Sciences(HICSS-34)*, pp. 1007, Wailea, HI, 3-6 January, 2001

[20] P. Doherty, "Learner control in asynchronous learning environments" *ALN Magazine*, vol. 2, 1998.

[21] D. Alex, C. Michael, and A. Peter, *Risk management for software projects*, (2nd .ed), Berkshire, England: McGraw-Hill Book Company, 1994.

[22] G. Crisp, "A model for the implementation and sustainability of a course management system in a research university." in proceedings of *19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, pp. 129-138, Auckland, New Zealand:, 2002

[23] U.-D. Ehlers, L. Goertz, B. Hildebrandt, and J. M. Pawlowski, 2005,Quality in e-learning use and dissemination of quality approaches in european e-learning: A study by the european quality observatory *the European Quality Observatory*, [Online] Available:http://www2.trainingvillage.gr/etv/publication/download/panorama/5162_en.pdf.

[24] P. Henry, "E-learning technology, content and services" *Education and Training*, vol. 43, pp. 249-255, 2001.

[25] K. Conner and M. Conner, 2006,Elerning, *Ageless Learner*, [Online] Available:http://agelesslearner.com/intros/elearning.html.

[26] T. Heift and D. Nicholson, "Web delivery of adaptive and interactive language tutoring " *International Journal of Artificial Intelligence in Education*, vol. 12, pp. 310-324, 2001.

[27] A. Kobsa, "Generic user modeling systems" *User Modelling and User-Adapted Interaction Journal*, vol. 11, pp. 49-63, 2001.

[28] F. Esposito, O. Licchelli, and G. Semeraro, "Discovering student models in e-learning systems" *Journal of Universal Computer Science*, vol. 10, pp. 47-57, 2004.

[29] M. I. Santally and A. Senteni, "Adaptation models for personalisation in web-based learning environments" *Malaysian Online Journal of Instructional Technology*, vol. 2, 2005.

[30] G. Richter, "On the relationship between information and data " in proceedings of *5th Informatik Symposium*, pp. 21-43, Heidelberg, Berlin 1976

[31] A. P. S. Alan F Karr, David L Banks "Data quality: A statistical perspective" *Statistical Methodology* vol. 3, pp. 137-173 2006.

[32] R. Y. Wang, "A product perspective on total data quality management" *Communication of the ACM*, vol.41, pp. 58-65, February 1998.

[33] S.-A. Knight, *User perception of information quality in world wid web information retrieval behaviour*, PhD Thesis, Edith Cowan University.2008

[34] J. Juran., *The quality control handbook*, (3rd .ed), New York,: McGraw-Hill, 1974.

[35] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers" *Journal of Management Information Systems*, vol. 12, pp. 5, 1996.

[36] D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data quality in context" *Communications of the ACM*, vol. 40, pp. 103-110, 1997.

[37] G. Shankar and S. Watts, "A relevant, believable approach for data quality assessment" in proceedings of *8th international conference on information quality,* pp. 178-189, MIT, 2003

[38] B. D. Klein, "User perceptions of data quality:Internet and traditional text" *Journal of Computer Information Systems*, vol. 4, pp. 9-18, 2001.

[39] M. Porter, "Towards a dynamic theory of strategy" *Strategic Management Journal,* vol. 12, pp. 954-1117, 1991.

[40] I. Elpez and D. Fink, "Information systems success in the public sector: Stakeholders' perspectives and emerging alignment model" *Issues in Informing Science and Information Technology*, vol. 3, pp.219-231, 2006.

[41] W. H. Delone and E. R. Mclean, "Information systems success: The quest for the dependent variable." *Information Systems Research*, vol. 3, pp. 60-95, 1992.

[42] W. H. Delone and E. R. Mclean, "The delone and mclean model for information systems success:A ten year update." *Journal of Managment Information Systems.* , vol. 19, pp. 9-30, 2003.

[43] S. Lin, J. Gao, A. Koronios, and V. Chanana, "Developing a data quality framework for asset management in engineering organisations" *International Journal of Information Quality*, vol. 1, pp. 100 - 126 2007.

[44] P. Katerattanakul and K. Siau, "Measuring information quality of web sites: Development of an instrument" in proceedings of *20th international conference on Information Systems,* pp. 279-285, Charlotte, North Carolina, USA, 1999

[45] B. D. Klein, "When do users detect information quality problems on the world wide web?" in proceedings of *8th American Conference in Information Systems,* pp. 1101-1103, 2002

[46] S. A. Knight and J. Burn, "Developing a framework for assessing information quality on the world wide web" *Informing Science Journal*, vol. 8, pp. 195-172, 2005.

[47] M. Gertz, "Managing data quality and integrity in federated databases" in proceedings of *2nd Working Conference on Integrity and Internal Control in Information Systems,* pp. 211–230, 1996

[48] T. Redman, *Data quality for the information age,* (2nd .ed), London: Artech House, 1996.

[49] R. H. J. Zeist and P. R. H. Hendriks, "Specifying software quality with the extended iso model" *Software Quality Journal*, vol. 5, pp. 273-284, 1996.

[50] M. Jarke and Y. Vassiliou, "Data warehouse quality: A review of the dwq project" in proceedings of *2nd Intl. Conf. on Information Quality* pp. 98--112, Cambridge, Mass., 1997

[51] Y. Chen, Q. Zhu, and N. Wang, "Query processing with quality control in the world wide web" *World Wide Web Journal*, 1, 241–255, 1998.

[52] J. E. Alexander and M. A. Tate, *Web wisdom: How to evaluate and create information quality on the web* (4th .ed), New Jersey Lawrence Erlbaum Associates, 1999.

[53] G. Shanks and B. Corbitt, "Understanding data quality: Social and cultural aspects" in proceedings of *10th Australasian Conference on Information Systems;,* pp. 785, Victoria University of Wellington, 1-3 December, 1999

[54] A. Dedeke, "A conceptual framework for developing quality measures for information systems" in proceedings of *5th International Conference on information Quality,* pp. 126-128, MIT, USA, 2000

[55] X. Zhu and S. Gauch, "Incorporating quality metrics in centralized/ distributed information retrieval on world wide web" in proceedings of *23 international ACM*

SIGIR conference on research and development in information retrieval, pp. 288 - 295 Athens, Greece, 2000
[56] H. K. N. Leung, "Quality metrics for intranet applications" *Information & Management*, vol. 38, pp. 137 – 152., 2001.
[57] M. Eppler and P. Muenzenmayer, "Measuring information quality in the web context: A survey of state-of-the-art instruments and an application methodology." in proceedings of *International Conference on Information Quality*, pp. 187–196, Boston, MA, 2002
[58] K. Kahn, D. Strong, and R. Wang, "Information quality benchmarks: Product and service performance" *Communications of the ACM*, vol. 45, pp. 184 – 193, 2002.
[59] M. Mecella, M. Scannapieco, A. Virgillito, R. Baldoni, T. Catarci, and C. Batini, "Managing data quality in cooperative information systems" in proceedings of *Confederated International Conferences DOA, CoopIS and ODBASE*, pp. 486–502, California, USA, 2002
[60] X. W. Liu and S. L. Han, "Ranking fuzzy numbers with preference weighting function expectations" *Computers & Mathematics with Applications*, vol. 49, pp. 1731-1753, Jun, 2005.
[61] S. Besiki, L. Gasser, M. B. Twidale, and L. C. Smith, "A framework for information quality assessment" *Journal of the American Society for Information Science and Technology* vol. 58, pp. 1720-1733, 2007.
[62] K. Kimberly, M. Pankaj, and W. John, "Do you know your iq?: A research agenda for information quality in systems" *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 37, pp. 26-31, 2009.
[63] M. J. Kargar and F. Azimzadeh, "A framework for ranking quality of information on weblog" *World Academy of Science, Engineering and Technology*, vol. 56, pp. 690-695, 2009.
[64] T. T. P. Thi and M. Helfert, "An information system quality framework based on information system architectures " chapter in: *Information systems development: Challenges in practice, theory, and education*, (1st .ed).
Springer US, 2009, pp. 337-350.
[65] J. Han and M. Kamber, *Data mining concepts and techniques.*, (2nd .ed), Fransisco, USA: Morgan Kaufmann, 2006.
[66] P. Maes, "Agent that reduce work and information overload" *Communication of the ACM*, vol. 37, pp. 30-40, 1994.
[67] X. Zhou, Y. Li, P. Bruza, S. T. Wu, and Y. Xu, "Using information filtering in web data mining process" in proceedings of *ACM International Conference on Web Intelligence*, pp. 163-169, Silicon Valley, California, USA, Nov. 2-5, 2007
[68] G. Stumme, A. Hotho, and B. Berendt, "Semantic web mining: State of the art and future directions" *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, pp. 124-143, 2006.
[69] Leeds Metropolitan University, 2010,Finding and evaluating information on the web, [Online] Available:libraryonline.leedsmet.ac.uk/lco/publications/pdf/web/qg-46.pdf
[70] J. Sivaramakrishnan and V. Balakrishnan, "Web mining functions in an academic search application" *Informatica Economica Journal*, vol. 13, pp. 132-139, 2009.
[71] R. Kosala and H. Blockeel, "Web mining research: A survey" *ACM SIGKDD Explorations,*, vol. 2, pp. 1-15, 2000.
[72] G. Lappas, "An overview of web mining in societal benefit areas" *Online Information Review*, vol. 32, pp. 179-195, 2008.

[73] R. Cooley, J. Srivastava, and B. Mobasher, "Web mining: Information and pattern discovery on the world wide web" in proceedings of *9th International Conference on Tools with Artificial Intelligence* pp. 12-23, Newport Beach, CA 3-8 November, 1997

[74] R. Kosala and H. Blockeel, "Web mining research: A survey" *SIGKDD Explorations*, 2, 1-15, 2000.

[75] C. Siefkes, "Learning to extract information for the semantic web" in proceedings of *Berliner XML Tage* pp. 452-459, Berlin, 13-15 October, 2003

[76] A. N. Oppenheim, "Analytic survey designs" chapter in: *Questionnaire design, interviewing and attittude measurement*, (New .ed).
London: Pinter Publishers Ltd, 1993, pp. 21-37.

[77] S. Arber, "Designing samples" chapter in: *Researching social life*, N. Gilbert, (1st .ed).
London: SAGE Publications Ltd, 1993, pp. 68-92.

[78] R. Likert, "A technique for the measurement of attitudes" *Archives of Psychology* vol. 140, pp. 1–55, 1932.

[79] M. Procter, "Measuring attitudes" chapter in: *Researching social life*, N. Gilbert, (1st .ed).
London: SAGE Publications Ltd, 1993, pp. 117-134.

[80] B. Gillham, *Developing a questionnaire,* (1st .ed), london: Countinuum, 2000.

[81] A. N. Oppenheim, "Pilot work" chapter in: *Questionnaire design, interviewing and attittude measurement*, (New .ed).
London: Pinter Publishers Ltd, 1993, pp. 47-64.

[82] J. Carter, K. Ala-Mutka, U. Fuller, M. Dick, J. English, W. Fone, and J. Sheard, "How shall we assess this?" in proceedings of *Annual Joint Conference Integrating Technology into Computer Science Education,* 107–123, New York, NY, USA, 2003

[83] L. J. Cronbach, "Coefficient alpha and the internal structure of tests" *Psychometrika,* vol. 31, 1951.

[84] L. J. Cronbach, "Test validation" chapter in: *Educational measuremnt, american council on education,* (R.L. Thorndike .ed).
Washington DC: 1971, pp. 443-507.

[85] U. Sekaran, *Research methods for business: A skill-building approach*, (3rd .ed), New York: John Wiley and Sons., 2000.

[86] J. Pallant, *Spss survival manual*, (1st .ed), United Kingdom: Open University Press, 2005.

[87] J. P. Peter, "Reliability: A review of psychometric basics and recent marketing practices" *Journal of Marketing Research*, vol. 16, pp. 6-7, February, 1979.

[88] P. Robinson, R. Shaver, and S. Wrightsman, "Criteria for scale selection and evaluation" chapter in: *Measures of personality and social psychological attitudes,* (1st .ed).
San Diego, CA: Academic Press., 1991, 1-16.

[89] J. Churchill and A. Gilbert, "A paradigm for developing better measures of marketing constructs " *Journal of Marketing Research*, vol. 16, pp. 64-73, 1979.

[90] J. F. Hair, R. L. Tatham, R. E. Anderson , and W. Black, *Multivariate data analysis : With readings* (5th .ed), London: Pearson Education, 1998.

[91] A. Field, *Discovering statistics using spss for windows,* (1st .ed), Sage Publications Ltd, 2000.

[92] G. Hutcheson and N. Sofroniou, *The multivariate social scientist,* (1st .ed), London: Sage Publications Ltd, 1999.

[93] A. L. Comrey and H. B. Lee, *A first course in factor analysis,* (2nd .ed), Lawrence Erlbaum Associates Inc, 1992.

[94] B. Holmes, *Cross-cultural differences of use of information technology in education: A comparative study of the use of computers in japanese and british classrooms,* PhD thesis, Wolfson College, University of Cambridge,.1998

[95] C. M. Allwood and Z.-M. Wang, "Conceptions of computers among students in china and sweden." *Computers in Human Behavior,* 6, 185-99 1990.

[96] M. Brosnan and W. Lee, "A cross-cultural comparison of gender differences in computer attitudes and anxieties: The united kingdom and hong kong" *Computers in Human Behavior* 14, 559-577 1998.

[97] V. Makrakis, "Cross-cultural comparison of gender differences in attitude towards computers in japan and sweden " *Scandinavian Journal of Educational Research,* 36, 275 - 287 1992.

[98] E. Whitley Jr. Bernard, "Gender differences in computer-related attitudes and behavior: A meta-analysis" *Computers in Human Behavior*, 13, 1-22 1997.

[99] M. H. Omar, "Attitudes of college students towards computers: A comparative study in the united states and the middle east. " *Computers in Human Behavior*, 8, 249-257, 1992.

[100] F. Al-Harby, R. Qahwaji, and M. Kamala, "The effects of gender differences in the acceptance of biometrics authentication systems within online transaction." in proceedings of *Cyper World (CW 2009),* pp 203-210, Bradford, UK, September 07-September 11, 2009

[101] J. Cohen, *Statistical power analysis for the behavioral sciences,* (2nd .ed), Psychology Press, 1988.

[102] A. Field, *Discovering statistics using spss,* (3rd .ed), London: SAGE Publications Ltd 2009.

[103] E. R. Babbie, *The practice of social research,* (11th .ed), Belmont, Calif Wadsworth Pub, 1997.

[104] R. B. Kline, *Principles and practice of structural equation modeling,* (2nd .ed), New York: Guilford Press, 2004.

[105] R. P. Mcdonald, "A simple comprehensive model for the analysis of covariance structures" *British Journal of Mathematical and Statistical Psychology*, vol. 37, pp. 234-251, 1978.

[106] B. Byrne, *A primer of lisrel : Basic applications and programming for confirmatory factor analytic models,* (1st .ed), New York: New York: Springer., 1989.

[107] T. A.Brown, *Confirmatory factor analysis for applied research,* (1st .ed), New York: Guilford Press., 2006.

[108] L. T. Hu and P. M. Bentler, "Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. " *STRUCT EQU MODELING* vol. 6, pp. 1-55, 1999.

[109] F. Alfonso, L. Luigi, M. Sandro, C. Stefano, O. Giandomenico, and O. Elena, "Applying gqm in an industrial software factory" *ACM Transactions on Software Engineering and Methodology (TOSEM)* vol. 7, pp. 411-448, 1998.

[110] C. B. Lionel, M. D. Christiane, and H. D. Rombach, "Practical guidelines for measurement-based process improvement" *Software Process: Improvement and Practice*, vol. 22, pp. 253-280, 1996.

[111] V. R. Basili and D. M. Weiss, "A methodology for collecting valid software engineering data" *IEEE Trans. Software Eng.*, vol. 10, pp. 728-738, 1984.

[112] V. Basili, G. Caldiera, and D. H. Rombach, 1994, *The goal question metric approach* [Online]. Available: https://docweb.lrz-muenchen.de/cgi-bin/doc/nph-webdoc.cgi/000110A/http/scholar.google.de/scholar=3fhl=3dde&lr=3d&cluster=3d4068380033007143449

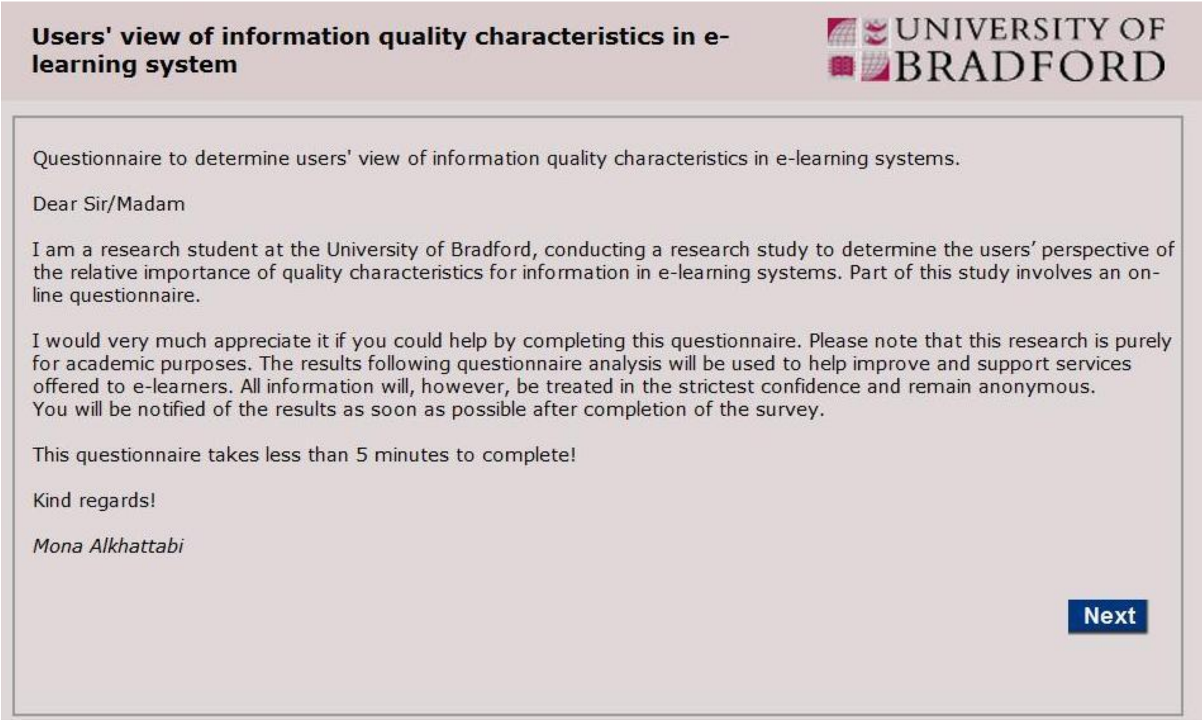
[113] N. Chen, B. H. Far, and Y. Wang, "Development of an intelligent agent-based gqm software measurement system" in proceedings of *first international conference on agent based technologies and systems (ATS2003)*, pp. 188-197, Canada, August, 2003
[114] F. Van Latum, R. Van Solingen, M. Oivo, B. Hoisl, D. Rombach, and G. Ruhe, "Adopting gqm-based measurement in an industrial environmental" *IEEE Software*, vol. 15, pp. 78-86, 1998.
[115] N. Khamis, S. Idris, and R. Ahmad, "Applying gqm approach towards development of criterion-referenced assessment model for oo programming courses" *WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY* vol. 32, pp. 152-156, August, 2007.
[116] Y. Jin, B. W. Choi, and J. Jerry, "An introduction of gqm approach for measuring educational activities for accreditation" in proceedings of *ICEE & ICEER*, Seoul, Korea, 23-28 August, 2009
[117] T. Gilb, *Software metrics*, (2nd .ed), Cambridge, Mass. : Winthrop Publishers, 1977.
[118] K. Buvaneswari, W. A. Venkataraman, and J. Ward, *"An introduction to software quality"* US Army Corps of Engineers, Waterways Experiment Station (WES), Washington, DC, Technical Report, no. ITL-99-4, June, 1999.
[119] K. Kahn and M. Ramakrishnan, "Assessment model for software maintenance tools: A conceptual framework" in proceedings of *Pasific Asia conference on Information systems (PACIS)*, pp. 527-536, Brisbane, Australia, 1997
[120] K. M. Khan and J. Han, "Assessing security properties of software components: A software engineer's perspective" in proceedings of *IEEE Australian Software Engineering Conference (ASWEC'06)*, pp. 199-208, 2006
[121] E. W. Weisstein. "About mathworld". [cited 2010 30th May]; Available from: http://mathworld.wolfram.com/about/.
[122] N. Egge and A. Krowne. "The planetmath story ". [cited 2010 30th May]; Available from: http://planetmath.org/?op=about.
[123] J. E. Rubin, *Set theory for the mathematician* (1st .ed), San Francisco, California: Holden-Day, 1967.
[124] R. R. Stoll, *Set theory and logic*, (1st .ed), San Francisco: Freeman and Company, 1961.
[125] T. Loton, *Web content mining with java: Techniques for exploiting the world wide web*, (1st .ed), West Sussex, England: Johns Wiley & Son Ltd,, 2002.
[126] *"Search engine user behavior study"* iProspect, Research Study Report, 2006.
[127] M. Alkhattabi, D. Neagu, and A. Cullen, "Information quality in information systems: A critique of relevant literature" in proceedings of *4th Saudi International Conference* Manchester, 30-31 July 2010
[128] M. Alkhattabi, D. Neagu, and A. Cullen, "Information quality framework for e-learning systems" *Knowledge Management & E-Learning: An International Journal (KM&EL)*, in press.
[129] M. Alkhattabi, D. Neagu, and A. Cullen, "User perceptions of information quality in e-learning systems: A gender and cultural perspective" chapter in: *Globalization, technology diffusion and gender disparity: Social impacts of icts* R. Pande, T.V.D. Weide, and N. Flipsen, IGI Global 2011,
[130] M. Alkhattabi, D. Neagu, and A. Cullen, "Validating an information quality framework for e-learning systems: An empirical approach" in proceedings of *15th International Conference on Software Process Improvement Research, Education and Training, INSPIRE (e-Learning and Social Responsibility)*, 47-54, BCS, London, 31st of March 2010

[131] M. Alkhattabi, A. Cullen, and D. Neagu, "Information quality metrics in an e-learning context: Applying gqm approach" in proceedings of *7th International Conference on the Quality of Information and Communications Technology,* Oporto, Portugal 29 September to 2 October 2010 2010

[132] M. Alkhattabi, D. Neagu, and A. Cullen, Assessing information quality of e-learning systems, *"Computers in Human Behavior"* Elsevier, **Under Review**

Appendices

Appendix I



Users' view of information quality characteristics in e-learning system

UNIVERSITY OF BRADFORD

Questionnaire to determine users' view of information quality characteristics in e-learning systems.

Dear Sir/Madam

I am a research student at the University of Bradford, conducting a research study to determine the users' perspective of the relative importance of quality characteristics for information in e-learning systems. Part of this study involves an on-line questionnaire.

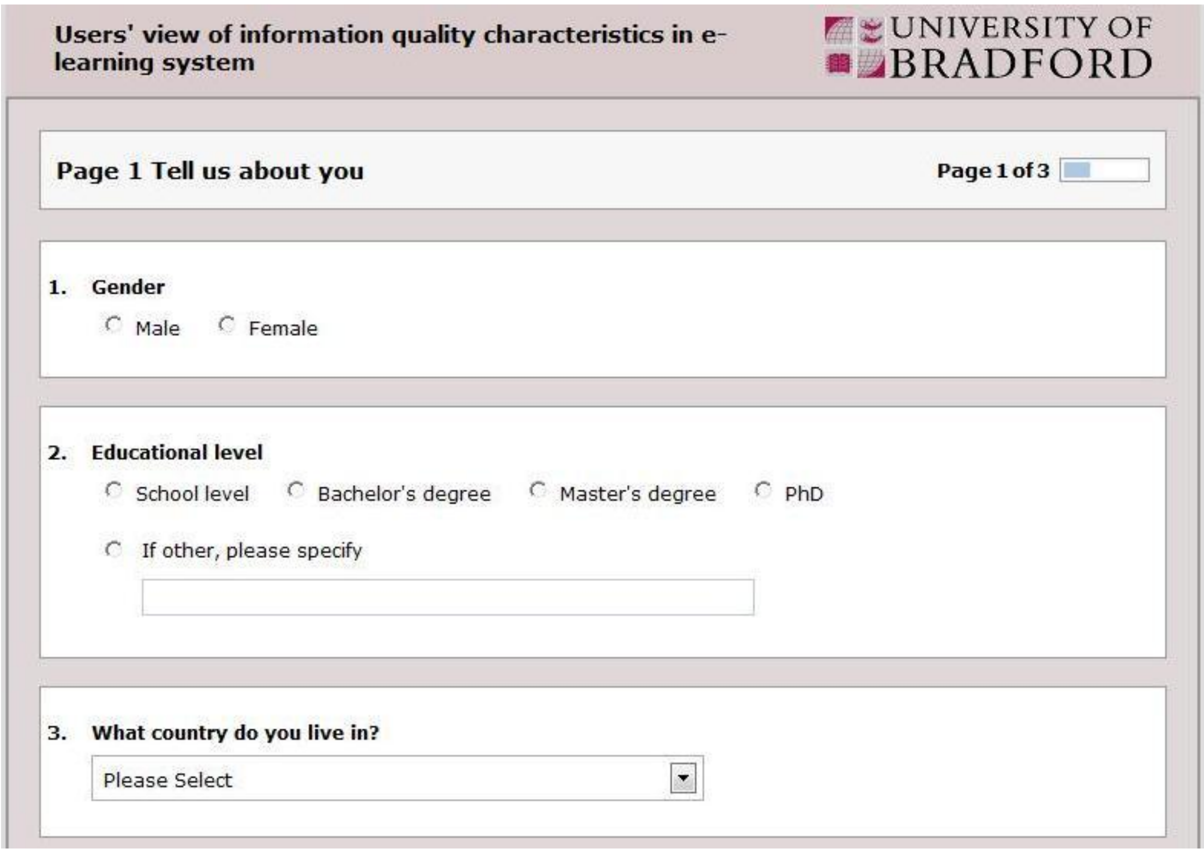
I would very much appreciate it if you could help by completing this questionnaire. Please note that this research is purely for academic purposes. The results following questionnaire analysis will be used to help improve and support services offered to e-learners. All information will, however, be treated in the strictest confidence and remain anonymous.
You will be notified of the results as soon as possible after completion of the survey.

This questionnaire takes less than 5 minutes to complete!

Kind regards!

Mona Alkhattabi

Next



Users' view of information quality characteristics in e-learning system

UNIVERSITY OF BRADFORD

Page 1 Tell us about you Page 1 of 3

1. Gender

Male Female

2. Educational level

School level Bachelor's degree Master's degree PhD

If other, please specify

3. What country do you live in?

Please Select

4. What kind of e-learning system(s) do you usually use? Y*__ou may select more than one option.__*

- ☐ Web supplemented (e.g. course outline and lecture notes online, use of email, links to external online resources etc)
- ☐ Web dependent (using the internet for main active elements in the learning procedure without any considerable reduction in face-to-face learning time such as online discussion, assessment and projects)
- ☐ Mixed mode (taking part in online activities, e.g. online discussions, assessment, online project, as part of the course work, which replace part of the typical classroom activities)
- ☐ Fully online(no physical interaction with classmates and instructors)
- ☐ Not sure
- ☐ Non
- ☐ If other, please specify

Next

Users' view of information quality characteristics in e-learning system

UNIVERSITY OF BRADFORD

Page 2 Your use of e-learning systems

Page 2 of 3

5. How many times do you access online learning systems weekly?

○ 0-5 ○ 6-10 ○ 11-20 ○ 21-30 ○ 31+

6. Do you use e-learning systems as

○ Learner/Student ○ Author/Teacher

○ If other, please specify

Previous Next

Users' view of information quality characteristics in e-learning system

UNIVERSITY OF BRADFORD

Page 3 How important each of the following quality characteristics for you **Page 3 of 3**

7. How important each of the following quality characteristics for you

	Very Important	Important	Neutral	Not Important	Not at All Important
Providing believable information	○	○	○	○	○
Providing accurate information	○	○	○	○	○
Providing unbiased, unprejudiced and impartial information	○	○	○	○	○
The system has an overall quality as seen or judged by people in general	○	○	○	○	○
The provided information is consistent and does not contain any conflict in meaning	○	○	○	○	○
The available materials contain value-added components such as interactivity and multimedia	○	○	○	○	○
Available information is applicable and helpful for your task	○	○	○	○	○
The available information is up-to-date	○	○	○	○	○
The available resources have all the needed information	○	○	○	○	○
The volume of provided information is appropriate for your task	○	○	○	○	○
The provided information can be easily checked for correctness	○	○	○	○	○
Available information is in appropriate languages, symbols, and units, and the definitions are clear	○	○	○	○	○
The distributed information can be easily understood	○	○	○	○	○
The representation of information is consistent	○	○	○	○	○
Available information is concise	○	○	○	○	○
The information is accessible to all users and can be easily and quickly retrieved	○	○	○	○	○

Access to the available information is secure	◌	◌	◌	◌	◌
The system is quick to responed	◌	◌	◌	◌	◌
Information resources in the system are available for use whenever needed	◌	◌	◌	◌	◌

8. Finally, do you have any further comments

500

Previous Submit

Appendix II

Anti-image Correlation Matrix

	Believability	*Accuracy*	*Reputation*	*Consistency*	*Relevancy*	*Objectivity*	*Completeness*	*Amount Of Information*	*Verifiability*	*Interpretability*	*Understandability*	*Representational Consistency*	*Conciseness*	*Accessibility*	*Response Time*	*Availability*
Believability	**.828a**	-.252	.035	.114	-.105	-.231	-.121	.064	-.142	-.170	.100	-.159	.003	.055	.100	-.027
Accuracy	-.252	**.874a**	.038	-.117	-.159	-.237	-.076	.077	.071	-.107	-.114	.005	.015	.035	-.085	.014
Reputation	.035	.038	**.834a**	-.238	.041	-.243	-.270	.152	-.253	-.038	-.039	.002	-.158	.133	.018	.074
Consistency	.114	-.117	-.238	**.868a**	-.171	-.072	.001	-.103	.206	-.148	-.163	-.202	.067	-.148	.000	-.055
Relevancy	-.105	-.159	.041	-.171	**.890a**	-.040	-.109	-.162	-.025	.109	.047	-.029	.092	-.121	-.027	-.169
Objectivity	-.231	-.237	-.243	-.072	-.040	**.838a**	.086	-.152	-.011	.049	-.015	-.035	.129	-.006	-.076	.061
Completeness	-.121	-.076	-.270	.001	-.109	.086	**.881a**	-.247	.010	.065	-.044	.089	-.150	-.152	-.181	.091
AmountOfInformation	.064	.077	.152	-.103	-.162	-.152	-.247	**.861a**	-.357	-.107	.017	.011	-.149	-.043	.109	.027
Verifiability	-.142	.071	-.253	.206	-.025	-.011	.010	-.357	**.876a**	-.160	-.167	-.012	-.070	-.141	-.090	-.039
Interpretability	-.170	-.107	-.038	-.148	.109	.049	.065	-.107	-.160	**.921a**	-.077	.015	-.094	-.029	-.179	-.168
Understandability	.100	-.114	-.039	-.163	.047	-.015	-.044	.017	-.167	-.077	**.934a**	-.091	-.231	-.042	-.054	-.096
RepresentationalConsistency	-.159	.005	.002	-.202	-.029	-.035	.089	.011	-.012	.015	-.091	**.876a**	-.381	-.018	-.140	.074
Conciseness	.003	.015	-.158	.067	.092	.129	-.150	-.149	-.070	-.094	-.231	-.381	**.865a**	-.036	.122	-.120
Accessibility	.055	.035	.133	-.148	-.121	-.006	-.152	-.043	-.141	-.029	-.042	-.018	-.036	**.919a**	-.058	-.194
ResponseTime	.100	-.085	.018	.000	-.027	-.076	-.181	.109	-.090	-.179	-.054	-.140	.122	-.058	**.881a**	-.238
Availability	-.027	.014	.074	-.055	-.169	.061	.091	.027	-.039	-.168	-.096	.074	-.120	-.194	-.238	**.881a**

[a.] *Measures of Sampling Adequacy [MSA]*

Appendix III

Coefficients for the first factor [a]

	Correlations		
	Zero-order	Partial	Part
Believability	.388	.000	.000
Accuracy	.353	.000	.000
Objectivity	.368	.000	.000
Reputation	.672	1.000	.162
Relevancy	.393	.000	.000
Completeness	.716	1.000	.157
Amount Of Information	.708	1.000	.139
Verifiability	.764	1.000	.142
Understandability	.714	1.000	.133
Representational Consistency	.658	1.000	.156
Conciseness	.777	1.000	.140
Accessibility	.480	.000	.000
Response Time	.422	.000	.000
Availability	.396	.000	.000

[a] *Dependent Variable: First Factor*

Coefficients for the second factor [a]

	Correlations		
	Zero-order	Partial	Part
Believability	.284	1.000	.000
Accuracy	.376	-1.000	.000
Objectivity	.276	-1.000	.000
Reputation	.276	1.000	.000
Relevancy	.647	1.000	.237
Completeness	.463	-1.000	.000
Amount Of Information	.438	.987	.000
Verifiability	.494	-1.000	.000
Understandability	.486	-.992	.000
Representational Consistency	.386	.973	.000
Conciseness	.395	1.000	.000
Accessibility	.760	1.000	.306
Response Time	.732	1.000	.307
Availability	.758	1.000	.267

[a] *Dependent Variable: Second Factor*

Coefficients for the third factor [a]

	Correlations		
	Zero-order	**Partial**	**Part**
Believability	.768	1.000	.327
Accuracy	.733	1.000	.248
Objectivity	.849	1.000	.439
Reputation	.387	.000	.000
Relevancy	.394	.000	.000
Completeness	.351	.000	.000
Amount Of Information	.324	.000	.000
Verifiability	.370	.000	.000
Understandability	.316	.000	.000
Representational Consistency	.347	.000	.000
Conciseness	.238	.000	.000
Accessibility	.224	.000	.000
Response Time	.302	.000	.000
Availability	.212	.000	.000

[a] *Dependent Variable: Third Factor*

Appendix IV

UNIVERSITY OF BRADFORD
MAKING KNOWLEDGE WORK

Exit this survey

Information Quality

1. Information quality satisfaction survey

Dear Sir/Madam

I am a research student at the University of Bradford, conducting ,as part of my research, a satisfaction survey for online published learning materials

I would very much appreciate it if you could help by completing this questionnaire. Please note that this research is purely for academic purposes. The results following questionnaire analysis will be used to help improve and support services offered to e-learners. All information will, however, be treated in the strictest confidence and remain anonymous.

Kind regards!

Mona Alkhattabi

1 / 6

Next

UNIVERSITY OF BRADFORD
MAKING KNOWLEDGE WORK

Exit this survey

Information Quality

2. Personal information

Tell us about you please

1. Gender

- Male
- Female

2. Age

- less than 21
- 21-23
- more than 23

2 / 6

Prev Next

UNIVERSITY OF BRADFORD
MAKING KNOWLEDGE WORK

Exit this survey

Information Quality

3. describion

Will you please take a few minutes to read the information about "Set theory/Subsets" published in two different WebPages.

Then I need your answers for some questions about how satisfy you are with the quality of the information within the published materials

3 / 6

Prev Next

UNIVERSITY OF BRADFORD
MAKING KNOWLEDGE WORK™

Exit this survey

Information Quality

4. First resource

Please take a moment to read the information within the following link Math World Wolfram. Then answer the questions.

***1. For each statement, please assign a quality score ranged from 0 to 10, where higher value reflects a better quality score**

	0	1	2	3	4	5	6	7	8	9	10
Conciseness of information											
Verifiability of information (how easy could the information be checked for correctness)											
How consistent is the representation of information											
How understandable is the information											
Amount of provided information											
How do you consider the reputation of the publisher											
Completeness (does the resource provide all needed information)											
Information availability											
Information relevancy for the specified topic "subset"											
Information Accessibility											
How do you found webpage's response time											
The objectivity of the resources											
Information accuracy											
Information believability											

2. will you please assign an over all quality score for the published materials

0 1 2 3 4 5 6 7 8 9 10

4 / 6

Prev Next

UNIVERSITY OF BRADFORD
MAKING KNOWLEDGE WORK™

Exit this survey

Information Quality

5. Second resource

Please take a moment to read the information within the following link Planet Math. Then answer the questions.

***1. For each statement, please assign a quality score ranged from 0 to 10, where higher value reflects a better quality score**

	0	1	2	3	4	5	6	7	8	9	10
Conciseness of information											
Verifiability of information (how easy could the information be checked for correctness)											
How consistent is the representation of information											
How understandable is the information											
Amount of provided information											
How do you consider the reputation of the publisher											
Completeness (does the resource provide all needed information)											
Information availability											
Information relevancy for the specified topic "subset"											
Information Accessibility											

How do you found webpage's response time	○	○	○	○	○	○	○	○	○	○	○
The objectivity of the resources	○	○	○	○	○	○	○	○	○	○	○
Information accuracy	○	○	○	○	○	○	○	○	○	○	○
Information believability	○	○	○	○	○	○	○	○	○	○	○

2. will you please assign an over all quality score for the published materials

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

5 / 6

Prev Next

UNIVERSITY OF BRADFORD
MAKING KNOWLEDGE WORK™

Exit this survey

Information Quality

6. Thank you

Many thanks

I appreciate your time and cooperation

6 / 6

Prev Done

Appendix V

Quality package documentation[11]

Package Class **Tree** **Deprecated** **Index** **Help**

PREV NEXT **FRAMES** **NO FRAMES** **All Classes**

A B C G M O Q U V

A

Accessibility - Class in quality

This class calculates the accessibility score as a double value range between 0 and 1.

Accessibility() - Constructor for class quality.Accessibility

Accuracy - Class in quality

This class calculates the accuracy score as a double value range between 0 and 1.

Accuracy() - Constructor for class quality.Accuracy

AmountOfInformation - Class in quality

This class calculates the Amount Of Information score, as a double value range between 0 and 1.

AmountOfInformation() - Constructor for class quality.AmountOfInformation

---

B

[11] The full documentation for the quality package could be accessed from: http://www.elearningquality.com/documentation/quality/package-summary.html

This method takes a web page's URL as a parameter, then it invokes WebParserWrapper class to parse the content of the URL.

getAcces(String) - Method in class quality.Accessibility

This method takes a web page's URL as a parameter, then it invokes WebParserWrapper class to parse the content of this page.

getAOI(String, String[], String[]) - Method in class quality.AmountOfInformation

This method takes a web page's URL and two arrays of needed keywords and provides keywords as a parameter (The provided keywords are identified from the web page using Google Keywords Destiny tools http://googlerankings.com/ultimate_seo_tool.php), then it invokes WebParserWrapper class to parse the content of this page.

getBel(String) - Method in class quality.Believability

This method takes a web page's URL as a parameter, then it invokes WebParserWrapper class to parse the content of the URL.

getCom(String, String[]) - Method in class quality.Completeness

This method take a web page's URL and an array of keywords as parameter, then it invokes the WebParserWrapper class to parse the content of this page.

getCons(String, String[]) - Method in class quality.Conciseness

This method takes a web page's URL and an array of keywords as parameters, then it invokes WebParserWrapper class to parse the content of this page.

getConsis(String) - Method in class quality.Consist

This method takes a web page's URL as a parameter and identifies the domain's URL.

getFeatureLinks(Vector) - Method in class quality.Accessibility

This method steps through a set of hyperlinks, it calls on isLive() to check broken links.

getFeatureString(Vector, String) - Method in class quality.Accuracy

This method step, through a set of input strings, invokes the match() method from the Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.AmountOfInformation

This method step, through a set of input strings, invokes the match() method from the Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.Believability

This method step, through a set of input strings, invokes the match() method from the Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.Completeness

This method step, through a set of input strings, invokes the match() method from the Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.Conciseness

This method steps, through a set of input strings, invokes the match() method from Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.Objectivity

This method step, through a set of input strings, invokes the match() method from the Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.Understandability

This method step, through a set of input strings, invokes the match() method from the Operator class to compare these strings against a string.

getFeatureString(Vector, String) - Method in class quality.Verifiability

This method steps, through a set of input strings, invokes the match() method from Operator class to compare these strings against a string.

getFound(Vector, Vector) - Method in class quality.Conciseness

This method steps, through a set of input sentences, calls upon getFound() to identify links and text and then invoke the match() method from the Operator class to carry out wildcard comparisons between these sentences and a set of templates.

getLinks(Vector) - Method in class quality.Accessibility

This method picks out all hyperlinks from a vector of input sentences.

getLinks(Vector) - Method in class quality.Objectivity

This method extracts all hyperlinks from a set of sentences, checks the domain extensions and counts the links with unbiased links.

getObj(String) - Method in class quality.Objectivity

This method takes a web page's URL as a parameter, it checks the domain extension, then it invokes WebParserWrapper class to parse the content of the URL.

getUnder(String) - Method in class quality.Understandability

This method takes a web page's URL as a parameter, then it invokes the WebParserWrapper class to parse the content of this page.

getVer(String) - Method in class quality.Verifiability

This method takes a web page's URL as a parameter, then it invokes WebParserWrapper class to parse the content of this page.

M

main(String[]) - Static method in class quality.overAll

main method which instantiated the other ten classes in order to calculate the overall quality score note: Availability score (av) and response time score (rsp) were calculated using Mon.itor.us online web monitoring tool Available from the following link: http://www.mon.itor.us , the Relevancy score (re) was calcutaed using Free Monitor for Google Version 2.5.28.75 Available to download from: http://www.cleverstat.com/en/google-monitor-query.htm while the Reputation score (repu)was calculated using Page Rank Checker Available from the following link: http://www.prchecker.info/check_page_rank.php all these scores were entered manulay in this code

---

O

Objectivity - Class in quality

This class calculates the objectivity score as a double value range between 0 and 1.

Objectivity() - Constructor for class quality.Objectivity

overAll - Class in quality

This class computed the overall quality score along the proposed quality framework using the assigned relative importance values.

overAll() - Constructor for class quality.overAll

Q

quality - package quality

U

Understandability - Class in quality

This class calculates the Understandability score, which will be either 1 if the web page provides an explanation component or 0 if not.

Understandability() - Constructor for class quality.Understandability

---

V

Verifiability - Class in quality

This class calculates the Verifiability score, which will be either 1 if the web page provides links to verify the information or 0 if not.

Verifiability() - Constructor for class quality.Verifiability

---

A B C G M O Q U V

Package Class **Tree** **Deprecated** **Index** **Help**

PREV NEXT **FRAMES** **NO FRAMES** **All Classes**