Enhance the search process of educational material by automating quality assessment of OERs

Georgia Zarnomitrou

Technische Universiteit Delft

TUDelft Delft University of Technology

Challenge the future

Enhance the search process of educational material by automating quality assessment of OERs

by

Georgia Zarnomitrou

in partial fulfillment of the requirements for the degree of

Master of Science
in Computer Science (Software Technology track)

at the Delft University of Technology,
to be defended publicly on Wednesday October 16, 2019 at 10:00 A.M

Supervisors:	Assistant Prof. Dr. C. Lofi	
	Dr. Michiel de Jong	
	M.E Munnik Msc	
Thesis committee:	Prof. Dr. Marcus Specht,	TU Delft
	Assistant Prof. Dr. C. Lofi,	TU Delft
	Assistant Prof Dr. M. A. Zuniga Zamalloa,	TU Delft

This thesis was realized in collaboration with TU Delft's Library and is confidential and cannot be made public until October 15, 2019

Preface

This work is a thesis on the subject of *enhancing the search process of educational material by automating quality assessment of OERs.* This project's aim is to find a way to automatically assess the quality of educational material in order to enhance the ranking given to them by current search engines. The report you are currently reading describes the work conducted during several months, specifically between the months November 2018 and August 2019. It has been written to obtain the degree of Master of Science at the Delft University of Technology within the program of Computer Science.

When my supervisors and I first discussed my research topic I was intrigued by the notion of researching the subject of quality of educational material and designing a quality model capable of assessing it. As a master student I am marginally aware of the restrictions and difficulties educators are faced with when looking for appropriate, high quality resources to achieve various educational goals. Therefore I could understand the need for improvements of current search processes and see the potential of using automated quality assessments as a way to achieve it. After delving into existing literature to research and understand the topic of quality of educational material, conducting experiments to understand my target audience, extracting and processing the metadata of educational resources from two different repositories, designing my quality model, writing my own code to implement it and conducting experiments to test the different rerankings it offers, I am happy with and proud of the final result.

I would like to thank my supervisors Professor Christoph Lofi, Dr. Michiel De Jong and Msc Michiel Munnik for their invaluable help over the past few months. Your guidance and feedback were helpful in providing me new ideas and challenged me to better my work and myself. I appreciate the time you devoted to this project and I enjoyed working with you. I would also like to thank all the professors, teaching assistants and Phd students that took the time to participate in my experiments. Furthermore I thank my friends for all the great moments inside and outside the lecture halls. Lastly, a special note of gratitude goes out to my parents and sister for their moral and emotional support during my time in Delft.

I am glad you are taking the time to read my thesis and I wish you a pleasant reading.

Georgia Zarnomitrou
Delft, September 2019

Contents

1 Introduction **1**
1.1 Introduction . . . 1
1.2 Problem Description . . . 2

2 Literature Review **5**
2.1 Introduction . . . 5
2.2 Key Concepts . . . 6
2.3 Analysis of the concept of Quality . . . 6
2.3.1 Quality Definitions and Models . . . 6
2.3.2 Quality Assessment in various fields and contexts . . . 9
2.3.3 Quality Assessment of educational material . . . 11
2.4 Discussion . . . 14
2.5 Conclusions. . . . 15

3 From Theory to Practice **19**
3.1 Introduction . . . 19
3.2 Determine facets of quality . . . 19
3.3 Meta-analysis . . . 20
3.4 Resource processing . . . 21

4 User Group Analysis **23**
4.1 Purpose of Analysis . . . 23
4.2 Hypothesis . . . 23
4.3 Measurements . . . 23
4.4 Set up . . . 24
4.4.1 Participants . . . 24
4.4.2 Set up . . . 24
4.5 How was the analysis conducted? . . . 24
4.6 Results . . . 24
4.7 Interpretation of results . . . 26
4.8 Threats to validity . . . 27
4.9 Conclusions. . . . 28

5 Methodology **29**
5.1 Prerequisites . . . 29
5.2 Diagram of Quality Model. . . . 29
5.3 Conceptual Design of Quality Model . . . 30
5.3.1 Step 1 . . . 30
5.3.2 Step 2 . . . 31
5.3.3 Step 3 . . . 31
5.3.4 Step 4 . . . 34
5.3.5 Step 5 . . . 35
5.3.6 Step 6 . . . 36

6 Implementation **41**
6.1 Technical Details . . . 41
6.1.1 Search Engine . . . 41
6.1.2 Resources . . . 41
6.1.3 Tools used for implementation. . . . 42
6.2 Functionalities of the prototype tool . . . 42
6.3 Result . . . 44

7 Evaluation **47**
7.1 Purpose of Analysis . . . 47
7.2 Hypothesis . . . 47
7.3 Measurements . . . 47
7.4 Set up . . . 48
7.5 How was the experiment conducted? . . . 48
7.6 Results . . . 49
7.7 Interpretation of Results . . . 50
7.8 Threats to validity . . . 51
7.9 Conclusions . . . 51

8 Conclusion **53**
8.1 Limitations . . . 53
8.2 Conclusions . . . 54
8.3 Future Direction . . . 55

A Appendix **57**

B Appendix **59**
B.1 Questionnaire for interviews . . . 59

C Appendix **61**
C.1 List of available query keywords for evaluation . . . 61

Bibliography **67**

1

INTRODUCTION

1.1. INTRODUCTION

Over the last two decades the internet has grown at an unprecedented rate leading to the creation and availability of a massive amount of digital information. Due to this growth in the late 20th century the Open Education Movement emerged, popularizing the idea that digital material can be designed to allow easy reuse in a wide range of teaching and learning situtations[1]. The idea for this movement was inspired by David Wiley's framework of openness known as the 5 Rs[2, 3]. According to this framework open educational resources (OERs) need to be freely accessed and shared so they can be reused, revised, remixed, redistributed and retained. The emergence of the Open Education Movement also prompted several universities like MIT, Harvard and Yale to provide free access to a substantial portion of their educational content. Having access to such a vast amount of educational resources offers great benefits to educators like flexibility, information and insights from other institutions, etc[4]. Furthermore, free access to all this open educational resources has led to a radical change in the way students learn, which in turn has affected the role of educators. In particular the role of educators is changing from that of sole providers of information to that of information managers. It is now up to educators to navigate through this sea of information and locate the most valuable content in order to fulfil their various tasks. The purpose of this thesis is to facilitate educators in finding the most useful, *high quality* open educational resources. Specifically, we want to improve the order in which educational materials appear in current search algorithms so that the most useful, high quality ones appear first.

Since the start of the Open Education Movement the main focus has been the creation of OERs for various subject areas[5] and building access to them by developing infrastructures, tools and repositories[6]. This has led to the availability of an enormous amount of OERs. Despite having access to such a vast amount of resources, educators are reluctant to use them because they do not trust or accept this new sharing cultures[6] and they perceive OERs as low **quality** material[5]. The reason for this negative view of OERs is probably caused by two of OERs' unresolved issues[7]: a) The discovery problem: Due to the large amount of resources available it is difficult and time-consuming for educators to sift through them and find exactly what they are looking for. b) The quality problem: From the vast amount of resources available it is difficult to find those of high quality. These issues are further exacerbated by the fast paced creation of OERs that does not allow for careful evaluation of their content and effectiveness. So, these issues hinder the use of OERs which could facilitate educators that consider creating their own material from scratch a difficult and labour-intensive task.

Taking all of this into account it appears that the web and open educational repositories offer such a large number of OERs that educators are having a hard time finding exactly what they want by simply querying search engines. Therefore to assist educators in better fulfilling their new role (as facilitators of knowledge) we need to enhance current search processes. Specifically, we need to think of more meaningful ways of ranking the results of a search query so that the most useful resources are easier to find. Considering the educators' claims regarding the lack of quality in OERs a good way of achieving this is by including a **quality assessment of the resources** in the ranking algorithm of a search engine. Recent changes in the educational landscape (growth of the internet, the rise of the Open Education Movement and appearance of Massive Open Online Courses) have led to the growing concern of the educational community about the topic of quality. They realize that establishing basic quality standards and models will improve the experience of students and educators alike[8]. A recent study [5] on the integration of OERs in course materials states the importance of developing cost-effective quality models and quality measures that will help educators find valid, reliable educational material in a time efficient manner. The creation of such a model is a way of promoting and realising the potential of OER to transform current educational practices, offer innovative forms for the creation and evaluation of OERs and provide empirical-based proof of their effectiveness[6]. So far a universally accepted quality model has not been developed due to the lack of comprehensive understanding of the complex nature of quality and the inability to reach a common consensus on its definition. All of this is evidence of the importance of quality and that an improved reordering of educational material can be achieved with the creation of the appropriate quality model.

In this thesis we will focus on gaining a comprehensive understanding of the concept of quality of educational resources in an effort to enhance current search processes by including quality measures in their ranking algorithm. Surprisingly the topic of quality of educational resources is mostly touched upon in a very peripheral way in the literature. Research so far has mostly focused on the quality of education as a whole, focusing on the way lessons are conducted or presented, the performance of Higher Education Institutions (HEIs), the quality of online or virtual education etc. Therefore to gain a thorough understanding of the meaning of quality for educational resources and to properly define it a literature survey on the topic is needed. From this survey we identify the various facets of quality that influence its definition and numerous criteria that contribute to the make up of the quality of OERs. Next we translate the findings of our literature review from a theoretical framework to a practical one so we can create a quality model to assist TU Delft educators. During that process we realize the need for a user group analysis that will give us insights in this group's needs and perceptions of quality. We also take the opportunity to corroborate the findings of our literature review through this user group analysis. Quality assessment of educational resources is currently achieved only through reviews given by recognized external parties, experts, peer reviews or assignment of ratings through tools. So, in order to automatically assess the quality of educational resources we compare the list of criteria that make up quality and the metadata attached to open educational resources. Then we design a quality model that uses these metadata to automatically assess the quality of OERs. After we design this quality model we implement and combine it with the results of an existing search engine to reorder its results in a more satisfactory way. Then experiments are conducted to ascertain if the re-ranking achieved with the automatic assessment of quality is an improvement over the existing ranking algorithm. Successfully completing these tasks means offering educators a way to more quickly and efficiently locate the educational material they need.

The main contributions of this thesis are as follows:

- We provide an extensive literature review on the topic of quality of Educational Resources. From this literature review we identify the factors that influence how quality of educational resources is perceived and a list of criteria used to assess it. Additionally, we offer our own definition for quality of educational material.
- We identify and present criteria that can be used to partially quantify quality in an automated way. This is achieved by comparing the list of criteria identified through literature and the available metadata information attached to open educational resources.
- We present a quality model that can be integrated in existing search engines. This model is capable of automatically calculating the quality of educational resources and reordering them based on it. More precisely, we offer four different mathematical formulas for the quantification of quality of OERs which result in four different re-rankings of search results.
- We present some initial experiment results regarding the satisfaction of educators with the various re-rankings of search results achieved through the quantification of quality. From this experiments we can ascertain if the inclusion of quality measures in ranking algorithms improves the ordering of the results and which formula achieves this more effectively.

1.2. PROBLEM DESCRIPTION

In this section we describe some key concepts that are needed to follow the work of this thesis, like information retrieval, search engines, relevance, retrieval models, metadata and open educational resources. Finally we finish this chapter by defining the main research questions of this work.

Information Retrieval

Since we are aiming to enhance the way current search engines work we first need to define the larger field of information retrieval. *Information retrieval* is the field concerned with the structure, analysis, organization, storage, searching and retrieval of information. [9]

Search Engines

In order to enhance existing search engines we first need to understand what they are and how they work. A *search engine* is the practical application of information retrieval techniques on large-scale text collections. More specifically, a *search engine* is the software system that compares queries to documents and produces ranked result lists of documents.[9]

Relevance

Relevance is a fundamental concept in information retrieval. Loosely speaking a relevant document contains the information that people are looking for when they submit a query to a search engine. Relevance can be distinguished in two categories *topical relevance* and *user relevance*.

Topical Relevance: A text document is considered to be topically relevant to a query if it is on the same topic.

User Relvance: A text document is considered to be user relevant if additional features of the document are taken into account to check if it is relevant for a specific user.[9]

Retrieval Model

The issue of relevance in search engines is addressed through retrieval models. A *retrieval model* is a formal representation of the process of matching a query and a document. *Retrieval models* form the basis for the ranking algorithms used in search engines to produce the ranked list of results. A good *retrieval model* uses both topical and user relevance to find documents matching a query. However some retrieval models focus only on topical relevance.[9]

Metadata

The term metadata was first used in 1968 by Philip Bagley, in his book “Extension of programming language concepts”. Metadata is defined as the data providing information about one or more aspects of other data, it is used to summarize basic information which can make tracking and working with specific data easier. The National Information Standards Organization (NISO), has separated metadata in three categories: descriptive (used for discovery and identification e.g. title, author etc.), structural (describe how components of object are organized) and administrative (provide information to help with managing the object e.g. file type, creation date)[10]. There are three metadata standards that are mostly used for educational resources and e-learning resources: Dublin Core, SCORM (Sharable Content Object Reference Model) and LOM (IEE Standard for Learning Object Metadata).

Open Educational Resources (OERs)

Open educational resources are educational materials which use a Creative Commons License (for more information on this topic look at Appendix A) or which exist in the public domain and are free of copyright. OERs is an overarching term that encompasses open textbooks, open courseware and other designations.[7]

Research Questions

Given the above definitions we specify the following research questions that need to be answered in order to achieve our goal of enhancing current search processes by automatically assessing the quality of OERs.

1. Gain understanding of the concept of quality of educational resources. This can be achieved by answering the following sub-questions:
   (a) What are the factors that influence the meaning of quality and how it is perceived?
   (b) What is the definition of quality of educational resources?
   (c) Is quality of educational resources quantifiable and to what degree?
   (d) What are the criteria that are used to assess the quality of educational resources?

2. Is it possible to quantify the quality of educational material automatically? To what extent and how? To answer this question we answer to the following sub-questions:
   (a) What information is required to quantify quality and what is actually available from the metadata of the resources?
   (b) What metric to use to achieve quality quantification?
3. Can the ordering of results given by current search algorithms be improved through the use of quality assessments of the educational materials?

To answer these questions we conduct an extensive literature review on the topic of quality of educational resources. Since there is little to no research that focuses exclusively on this topic we glean information on the quality of educational resources by research on quality of education in general. Additionally, we look at information from the research on quality in other fields and extract those that could be applied on educational resources. Then we put all these pieces of information together to gain a well-rounded understanding of quality of educational resources, identify the factors that influence how it is perceived and offer our own definition for it. From the literature review we also identify a list of criteria that could be used if someone wanted to quantify the quality of OERs. Using all this information we then look at the metadata available in open educational repositories and try to find if we can match them to the list of quality criteria. Specifically, we look at the metadata provided by three educational repositories MIT OpenCourseWare, TU Delft and OER Commons. Considering the limited information available in literature and in order to gain insights to the thought processes and preferences of our target group (meaning educators) we conduct a user group analysis. This user group analysis corroborates the findings of our literature review and offers some additional insights. Using our findings both from the literature review and the user group analysis we design a quality model that automatically assesses the quality of educational material, combines this assessment with the relevancy offered by the ranking algorithm of search engines and offers a new ranking of the search results. Then we implement this quality model and integrate it in a prototype search engine offered by the company Feedback Fruits. Finally, we conduct experiments to ascertain whether the new ranking offered by our quality model enhances the search process in a satisfactory way for our users. The metric in these experiments is the satisfaction of the users with the new orderings offered on a scale from 1 to 5 where one means "Not satisfied" and 5 means "Very satisfied". Initial experiments provide promising results regarding the automatic assessment of educational material and the re-ranking of search results based on their quality.

2

Literature Review

In this chapter we present our literature review on the topic of quality. From this review we build the framework for the rest of our work. Specifically, a critical analysis of the available literature allows us to gain a well-rounded view of the concept of quality and answer the first research question of this thesis.

2.1. Introduction

The purpose of this thesis is to help facilitate the task of finding high quality educational material for educators. We propose this could be achieved by automatically assessing the quality of educational material and using this assessment to enhance the way current search algorithms work. However in order to assess the quality of educational material a thorough understanding of the concept is needed. Attempting to use the notion of quality without a clear understanding of its complex nature entails the danger of it becoming a mere catchphrase, a word with high moral tone and little practical value. With this literature we gain the needed perspective regarding the concept of quality, and more precisely the quality of educational material which allows us to identify which factors influence it, identify the criteria used to assess it, determine whether it is quantifiable and define it. The importance of fully grasping the meaning of quality is further evidenced by the educational community's interest in this matter over the last two decades. Higher Education Institutions (HEIs) are exhibiting a noticeable concern over quality issues and numerous Quality Assurance (QA) organizations have been established across the world[11].

Using Google Scholar and the reference lists of any relevant articles we found, we searched for literature on the topic of quality of educational materials. Surprisingly, there has been very little research that focuses exclusively on the quality of educational material. The majority of the pertinent literature deals with this topic in a very peripheral way (the quality of the educational material is treated as a secondary parameter in these papers). Therefore we widened the scope of our search to include literature that deals with quality of education in general and even some papers that focus on quality in different fields of study. From this literature we abstract all the information and observations that could potentially be applied when thinking about the quality of educational resources. From our research we also observe that the majority of the literature focuses on ways to assure, improve or measure quality in a variety of contexts with little to no research on understanding this complex concept. This results in numerous papers and articles with scattered observations and suggestions of what is missing from quality assessments and how to improve them.

This literature review is structured as follows. In the second section we describe some key concepts regarding quality assessment in education like accreditation, benchmarking, certification etc. In the third section we present our findings, observations and deductions from the literature material. First we display the various definitions given to quality of education and some quality models that show how to achieve it. Next we present the deductions we made from research on the assessment of quality in different fields and contexts. After that we focus on the literature that focuses on the assessment of quality of educational materials. From these papers, first we present the snippets of information we glean from the papers that treat quality of educational resources in a peripheral way and then the few papers that focus exclusively on the topic. In the following section we summarize our findings and finally, we combine and synthesize all the acquired information to answer the first research question of this thesis.

2.2. KEY CONCEPTS

In this section I describe some of the key concepts related to the topic of quality assessment in education. These are concepts that keep appearing in the literature and show some of the ways in which quality assessments are managed in education, specifically how quality assessments are done in Higher Education Institutions.

CERTIFICATION:

Certification is interpreted as the level of recognition granted by the organization from which a quality standard model originated. Award of the certification will follow some form of review and may be accompanied by a requirement that the reviewed institution commits to an improvement plan and later renewal of certification[8].

BENCHMARKING:

Benchmarking is the process of comparison of institutional performance with that of others. Allocation to a benchmarking group indicates that either the originating organisation operates a benchmarking service or there is evidence of the model having been used in benchmarking exercises[8].

ACCREDITATION:

Accreditation is interpreted as a form of mandatory certification or licensing of institutions and/or their programmes that grants access to national financial support or recognition of awards for employment purposes[8].

QUALITY ASSURANCE

Quality assurance is a concept imported in education from the business sector. Quality Assurance (QA) is the process of establishing stakeholder confidence that provision (input, process and outcomes) fulfils expectations or measures up to threshold minimum requirements[12]. According to Cheng there are three different paradigms of QA in education [13]:

- Internal: Internal quality assurance focusses on improving the internal environment and processes so that the effectiveness of learning and teaching can be ensured to achieve the planned goals.
- Interface: Interface quality assurance is ensuring that education services satisfy the needs of stakeholders and is accountable to the public.
- Future: Future quality assurance focuses on ensuring the relevance of aims, content, practice and outcomes of education to the future of new generations.

QUALITY ENHANCEMENT

Quality Enhancement(QE) is a concept that is often associated and confused with quality assurance (QA). Quality enhancement is concerned with the improvement of learning and teaching processes. Quality assurance on the other hand emphasizes on the effectiveness of the educational process and evaluating the provision rather than the learning experience itself.

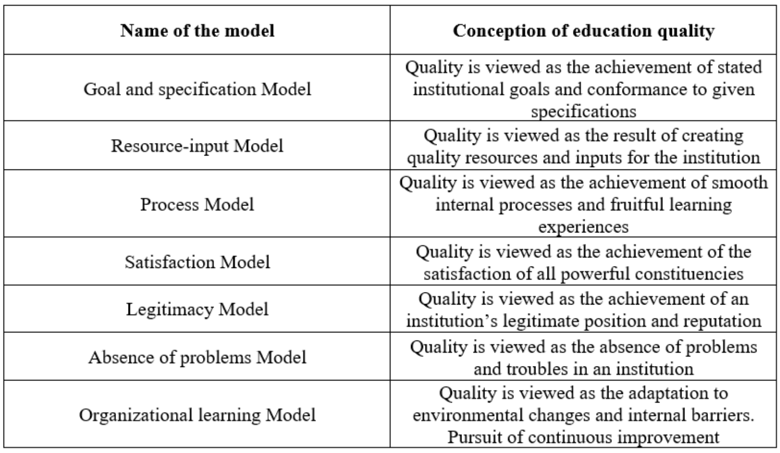
2.3. ANALYSIS OF THE CONCEPT OF QUALITY

In this section we provide a detailed analysis of the literature we found on the topic of quality. Specifically, we delve into the literature that deals with quality in education in general and get a generic overview of how quality is viewed and the existing strategies for achieving it. Then we look at how quality is assessed in specific contexts or other fields and extract any observations or methodologies that might prove useful when dealing with the quality of educational material. Lastly we examine the literature pertaining to the assessment of quality of educational materials.

2.3.1. QUALITY DEFINITIONS AND MODELS

While searching for material pertaining to the topic of quality we uncovered two papers that present quality definitions and models and one book that deals with quality in education. In this section we delve into the content of these materials and get a general idea of the current state of the art regarding quality in education.

In the early 1980s the concept of quality emerged in higher education from the industrial and business fields in which this topic had been well established[14]. By the end of the 1990s quality came to be seen as a quantifiable value and efforts were being made to define, achieve and measure it[11]. One of these efforts was the introduction of seven quality models that help form a framework to understand and conceptualize quality in education from different perspectives[15]. These models are presented in table 1.



Name of the model	Conception of education quality
Goal and specification Model	Quality is viewed as the achievement of stated institutional goals and conformance to given specifications
Resource-input Model	Quality is viewed as the result of creating quality resources and inputs for the institution
Process Model	Quality is viewed as the achievement of smooth internal processes and fruitful learning experiences
Satisfaction Model	Quality is viewed as the achievement of the satisfaction of all powerful constituencies
Legitimacy Model	Quality is viewed as the achievement of an institution's legitimate position and reputation
Absence of problems Model	Quality is viewed as the absence of problems and troubles in an institution
Organizational learning Model	Quality is viewed as the adaptation to environmental changes and internal barriers. Pursuit of continuous improvement

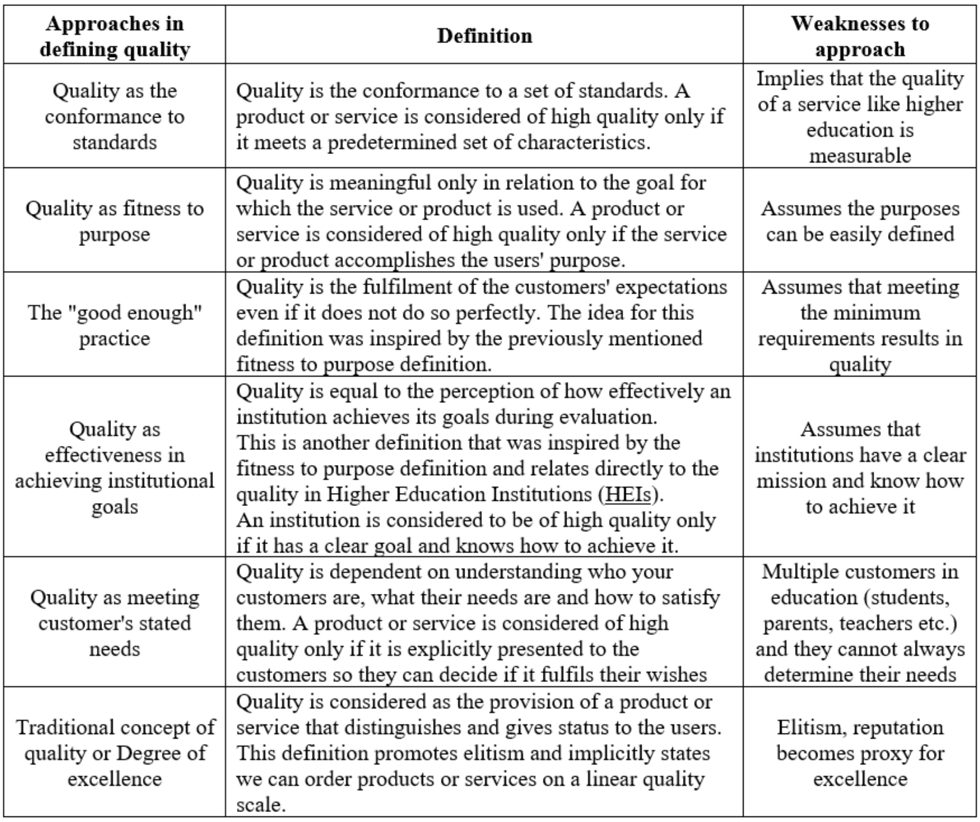
Table 1: Seven models of education quality

According to the authors of this work each model focuses on a different aspect of the educational process and pursues quality in that context. Additionally, they hypothesize that an institution that applies all seven models may achieve total quality. They consider that each model focuses on certain aspects of education quality, so their combination would provide a well rounded approach to achieving it. Besides the presentation of these models this work contains several observations regarding education quality. Mainly that "previous efforts aimed at improving education quality are suffering from poor understanding of its complex nature", "education quality is a vague and controversial concept" and that even though there are multiple definitions for quality which are highly correlated there is no universal agreement on its definition. One of the authors (Cheng) also offers his own definition of education quality. According to him, education quality is the character of the set of elements in the input, process, and output of the education system that provides services which completely satisfy both internal and external strategic constituencies by meeting their explicit and implicit expectations.

By observing the presented quality models we can agree that quality is a complex, multi-dimensional concept. Furthermore we can identify some aspects of quality, namely that it depends on the satisfaction of the person who judges it, the achievement of certain requirements and expectations, and it requires continuous improvement. Although the authors mention the need for continuous improvement in order to keep up with the changing educational environment, they fail to interpret this as quality's dependency on time. Additionally, they recognize different aspects of quality and the need to combine them, however they do not present one single model but rather seven different ones to be used in targeted ways. Lastly, they present these quality aspects as indicators of how to assess quality but do not recognize they are also influencing factors of how quality is perceived.

Nearly 20 years after the presentation of these models the educational community still struggles to define the concept of quality as evidenced by a paper written in 2015 that attempts to analyse the concepts of quality, Quality Assurance (QA) and Quality Enhancement(QE)[11]. In this work the author reiterates that quality is elusive and difficult to define, then proceeds by presenting and criticizing the six most-commonly used definitions of quality in the setting of higher education institutions. These definitions and their weaknesses as perceived by the author are displayed in table 2. After criticizing these definitions it is suggested that the difficulties in defining quality stem from two of its aspects, namely that it is a relative and context-dependent concept. Next an analysis of the concept of quality assurance and quality enhancement shows that quality also depends on the purpose and that it requires continuous improvement (see internal and future QA in section 2.2). The author then concludes that education quality is different than in other field areas. So she suggests that to properly define it a good understanding of the education process is needed and quality definitions should be studied by researchers from different backgrounds to gain perspective from many angles rather attempting to find a unified definition.

Reviewing the body of this work we agree that further research is needed on the topic of quality not necessarily to define it but mostly to fully grasp its essence. However we believe that quality should not be treated



Approaches in defining quality	Definition	Weaknesses to approach
Quality as the conformance to standards	Quality is the conformance to a set of standards. A product or service is considered of high quality only if it meets a predetermined set of characteristics.	Implies that the quality of a service like higher education is measurable
Quality as fitness to purpose	Quality is meaningful only in relation to the goal for which the service or product is used. A product or service is considered of high quality only if the service or product accomplishes the users' purpose.	Assumes the purposes can be easily defined
The "good enough" practice	Quality is the fulfilment of the customers' expectations even if it does not do so perfectly. The idea for this definition was inspired by the previously mentioned fitness to purpose definition.	Assumes that meeting the minimum requirements results in quality
Quality as effectiveness in achieving institutional goals	Quality is equal to the perception of how effectively an institution achieves its goals during evaluation. This is another definition that was inspired by the fitness to purpose definition and relates directly to the quality in Higher Education Institutions (HEIs). An institution is considered to be of high quality only if it has a clear goal and knows how to achieve it.	Assumes that institutions have a clear mission and know how to achieve it
Quality as meeting customer's stated needs	Quality is dependent on understanding who your customers are, what their needs are and how to satisfy them. A product or service is considered of high quality only if it is explicitly presented to the customers so they can decide if it fulfils their wishes	Multiple customers in education (students, parents, teachers etc.) and they cannot always determine their needs
Traditional concept of quality or Degree of excellence	Quality is considered as the provision of a product or service that distinguishes and gives status to the users. This definition promotes elitism and implicitly states we can order products or services on a linear quality scale.	Elitism, reputation becomes proxy for excellence

Table 2: Various definitions of quality in education and their weaknesses

separately just in the field of education but rather considered separately in the various contexts it appears in. Since quality is dependent on the context, the purpose and the perceptions of the target audience we consider it futile trying to define quality in general abstract terms with no point of reference. Quality should only be defined in reference to something concrete after careful consideration of all the parameters (context, purpose, target audience and their needs and expectations). As with the previous paper the author identifies the need for continuous development and improvement in education quality (see quality enhancement in section 2.2) but fails to see its dependency with time. Lastly, none of the above definitions offers a global well-rounded view on the concept of quality since each one disregards certain aspects of it or overemphasizes one aspect to the detriment of the rest.

The topic of quality and educational quality is also addressed in the book "Total Quality Management in Education"[16]. In this work findings on quality from the business arena are used and translated into the educational context. At the beginning some general remarks and observations are made, namely quality is described as "an enigmatic concept, perplexing to define and difficult to measure". Then the subjective nature of quality is mentioned and quality is equated with customer satisfaction and an aspiration for excellence. Quality is also described as an ephemeral concept that requires constant vigilance from those providing it if they are to keep meeting their customers' needs. Next the author suggests that quality is a dynamic idea that possesses an emotional and moral force and that is why it is difficult to define. He also cautions researchers against overanalysing it because it might lose its vitality. Additionally, it is inferred that the confusion over quality stems from the fact it can be used both as an absolute and a relative concept. Following is an explanation of these two views:

- **Quality as an absolute:** Quality as an absolute means striving for the highest possible standard (excellence), it is considered a rare commodity and it conveys prestige. In an educational context this can be translated as elitism. According to this view quality is an attribute that can be attached to a product or service.

- **Relative notion of quality**: Quality is viewed as a value that can be ascribed to a product or service and measured up against predetermined criteria time and time again. Based on this view quality is not an

> end in itself but rather a means by which a product or service is judged to see if it fits the purpose for which it is intended.

It is further suggested that the achievement of quality relies on understanding the needs and inclinations of the customers since they are the ones that end up judging the service or product. When trying to transfer these observations in an educational context the difficulty of the task is pointed out. Specifically, it is mentioned how hard it is to pinpoint the requirements of the customers since there is a variety of them in education. Finally it is argued that education is a service and service quality is more difficult to achieve and define because it is intangible, prone to subjective judgement and more difficult to define parameters for. Also it is mentioned that once a customer loses faith in a service it is harder to change their perception of its quality.

Observing the two differing views on quality described in this work, absolute and relative we object to the view of quality as an attribute that can be attached to a product or service because that would imply constancy. As we have already observed quality is a concept that depends on time and requires constant update and improvement. As for the relative notion of quality the idea that it needs to be constantly affirmed and tested is in line with the time dependency of quality. However, the author suggests that the requirements quality is tested against are constant. We refute this claim because of the individuality of customers. Based on the person who is judging the product or service, the context and the purpose of use, the requirements needed also change. This work focuses too much in the satisfaction of the customers as the means to achieving quality. Although customer satisfaction is one of the major factors in quality it should not be the only consideration when trying to achieve quality. Lastly, we agree that once the trust of customers is lost it is difficult to revert their opinion, which is additional proof that time is a factor in quality considerations, as a person accumulates experiences his/her beliefs and inclinations change.

2.3.2. Quality Assessment in various fields and contexts

In this section we present literature that deals with the concept of quality and more precisely quality assessment in a different field of study and in two different educational contexts. Specifically, we offer our findings from literature concerning the assessment of service quality, quality in virtual education and, online and open education.

Although service quality is in a different field than quality of educational material some information can still be gleaned from it and used for our purpose. So, we start with a paper in which 19 service quality models are critically appraised[17]. In the first and third models presented, quality is viewed as understanding the customers' expectations of quality service and meeting them in order to achieve consumer satisfaction. In the second model 5 dimensions of service quality are used: reliability, responsiveness, tangibles, assurance and empathy. From these dimensions three can also be applied to quality of educational resources, namely reliability, tangibles and empathy (we want the material to be reliable, it is something tangible and we need to know and understand the people who will use it). In another model, service quality is related not only to consumers' satisfaction but also their intentions. Therefore the purpose for which a service or product is needed also influences how quality is perceived. Another model views service quality as the satisfaction of predetermined attributes desired by the customers, meaning they have a preconceived notion of the ideal service or product. Expanding this idea another model presents quality as the comparison of customers' preconceived standard of excellence with their perception of the service offered. So customers could potentially settle for the service or product that best matches their view of excellence if it cannot be met. Another model suggests that customers' quality evaluations are based on a comparison between benefits and sacrifices. Therefore, when assessing quality between a multitude of services or products customers may end up sacrificing some of the desired attributes. Lastly, in another model, service quality is related to customers' post-purchase satisfaction and measured with their willingness to repurchase. This suggests that quality depends on customers' experiences.

In an overview of the state of quality in virtual education environments the author indicates the problems in the assessment of quality, which stem from the criteria used to measure it and advices the development of quality seals to improve the situation[18]. She also criticizes the way quality is dealt with in education in general, claiming it is a fragile term, empty of meaning defended based solely on academic achievements. Instead she suggests an alternative, namely to follow the constructivist principles[1] when dealing with quality. This suggestion implies that the experiences and knowledge accumulated over time affect people's perception

[1] Constructivism in education: framework in which learners construct new ideas and concepts based on their current and past knowledge

of quality. So, we hypothesize that the educational level of a person is also an influencing factor when defining quality. The author of this work continues by stating her vision of the future in which the web will use quality seals given by competent bodies to recognize and provide reliable and worthy content. From this we can infer that *reliability* and *"worthiness"* are two criteria used in the assessment of quality. Next the author presents a list of errors in the criteria used to assess quality of virtual education:

- The translation of insights regarding quality assessment directly from the business world. We agree with the author that this is wrong because the context and goals are different in an educational setting.
- The absolute reliance on user satisfaction for quality assessment. Again we agree with the author that only focusing on user satisfaction to assess such a complex concept as quality is short-sighted.
- The content used and support it offers are treated as if they are of less importance during assessment of multimedia systems. The limited literature concerning educational material corroborates this oversight.

The author concludes this work by offering her proposal on how to assess quality in virtual education based on 6 levels of approximation of the educational process. These levels are the following 1)educational action and scenario, 2) purposes of participants, 3) instructional agenda, 4) educational interaction and intervention, 5) educational instruments and 6)knowledge building. Looking at this proposal we see that in an effort to minimize the use of user satisfaction in quality assessment the author makes the same mistake she identified. Namely, her proposal is overly focused on assessing quality based on purpose and context dismissing the important factor of user satisfaction.

We conclude this section with a paper that shows the state of the art in quality models used for assessment in online and open education and offers recommendations for future improvements[8]. In this work it is stated that there is an extensive selection of quality models to select from based on institutional context, aim and maturity. So we see that the influencing factors for quality assessment are context, purpose and maturity. Next a list of characteristics that need to be addressed in quality systems of online education to achieve quality assurance and quality enhancement is presented:

- The multi-faceted nature of quality means that a multiplicity of measures are needed to get a well-rounded, holistic view of it.
- The systems have to be dynamic and flexible in order to adjust to the constantly changing environment (technological advancements and social perceptions).
- Quality systems have to be able to balance the perception and demands of various stakeholders.
- Quality systems have to be multi-functional, meaning to provide guide for future improvement and serve as a label for outside perspectives

The authors of this work acknowledge the merits of quality but also the difficulty in dealing with it because views of what constitutes quality are subjective and subject to purpose and context specific factors. They also recommend that cultural differences both geographic and institutional should be taken into account when analysing quality. Focusing on the quality in open and online education they claim it is an elusive and complex concept, something that is created or caused due to its context rather than something that constantly exists. This is in line with our previous observation that it is futile to try to define quality in an abstract way with no point of reference. Next the authors present the three domains touched upon by all quality standard models: a) Services (staff support, student support), b) Products (curriculum design, course design, course delivery), c)Management (strategic planning and development). Observing this domain it is clear that course materials are of secondary importance in quality assurance and the focus given to them minimal (evidenced by the focus on structure and design of the courses but not the actual material used). Then an explanation of the process of accreditation and certification is offered. Specifically, these processes involve peer review teams that manually go over documentation and interrogate staff in institutions to verify statements made by them in documentary submissions. A short list of criteria used in these quality assessments of documentation is offered, these criteria are *suitability, relevancy and appropriateness*. Focusing on quality assessment in MOOCs, the authors present the two indicators of quality used, namely *the reputation of the platform and the reputation of the institution based on its campus-based standing*. Finally the authors conclude by suggesting the creation of an e-learning resource hub where all educational material, quality tools and research papers could be gathered. They believe that the existence of such a hub would be useful both for institutions seeking to improve their practices and for the peer review teams responsible for quality assurance.

2.3.3. QUALITY ASSESSMENT OF EDUCATIONAL MATERIAL

In this section we offer our findings from reading papers that deal with quality of educational material or content. First we present the pieces of information we were able to abstract from the papers that deal with this topic in a peripheral way and then we delve into the ones that deal with it directly.

From a study about the quality and suitability of written educational materials for patients[19],we were able to extract a list of criteria used to assess their quality. Specifically, this study mentions the need to have educational material that is *readable, understandable, realistic, current and based on scientific foundations*. The results of this study provide another list of criteria used to assess quality of educational resources. Specifically the results showed that the educational materials that were evaluated were unsuitable due to their *content, structure, design, composition and language*. As far as the content is concerned patients claimed that it was complicated, difficult to understand and insufficient. However high points were awarded to cultural suitability of the materials. Therefore culture is another factor that influences people's perception of quality.

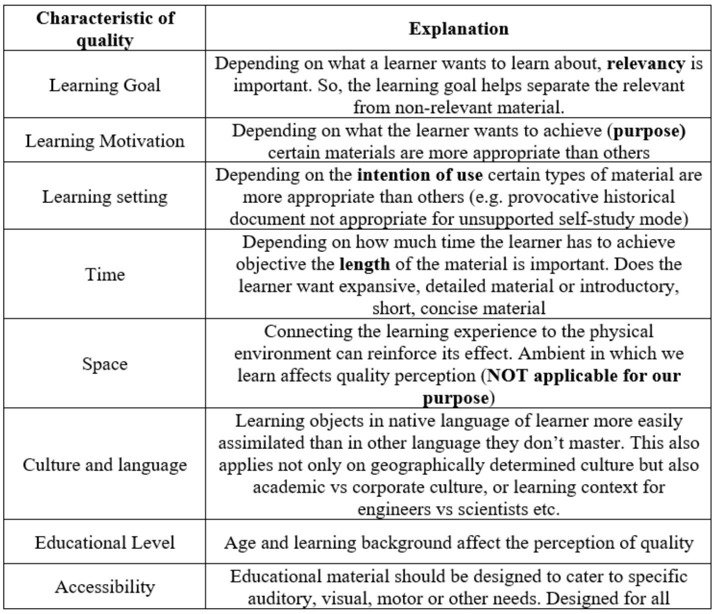
In a proposal for a new e-learning assessment model (the hexagonal e-learning model) quality is assessed based on 6 dimensions[20]. One of these dimensions deals with content quality and the parameters used to assess it check if the content is *well-organized, effectively presented, clearly written, of the right length, useful and up-to-date*. During a survey conducted to test this model, content validity was assessed based on the extent to which the measurements collected from the participants (students) reflect the specific intended domain of content, meaning that the primary criterion was the *relevancy* of the content.

Next we studied a paper that shows the results of an empirical study that investigated the relationship between user cognitive styles and perceptual multimedia quality[21]. From this work we were hoping to identify quality criteria that pertain to the assessment of multimedia educational content. The authors start by criticizing the parameters used to assess quality of multimedia information content because they do not convey needed information like the specific needs of application, informational load on user multimedia experience and do not encapsulate user media preferences (video over text) or characteristics. We agree with the authors' criticism and deduce that quality assessment of any information content should take into account the purpose for which the material is intended as well as the users' preferences. During the experiments of this study the multimedia was altered so that it contained losses, errors and its presentation (*resolution*) was degraded. The results were not definitive but measurements showed that multimedia presentation could be significantly degraded without proportional decrease in users' perception of quality. This suggests that the technical characteristics of multimedia information material plays a small role in how people perceive quality.

In a paper that shows the current approaches in the area of electronic documents for finding, reusing and adapting documents for teaching or learning purposes, one of the factors of these processes is finding high quality OERs[22]. Firstly, it is commented what a time consuming and labour intensive task it is to select the best documents and references from the large amount of available educational material due to the time it takes to assess them. Then the authors expound by mentioning how OERs are assessed, namely that *the qualifications of the authors are checked, recommendations are read, a cross-referencing with recognized digital academic libraries is conducted* and *copyright licenses are checked*. Next the authors express their displeasure with the lack of a universal meta-data standard for educational resources. It is then mentioned that the Dublin Core Educational Working Group has recognized the need for more metadata labels like "audience" (who is the intended audience of these resources), "conforms To"(the learning objectives), "Quality"(proof that an educational resource has been certified by some recognized body) etc. We agree that this is important information to have regarding educational resources and believe that their existence in the future could prove useful for the purpose of this thesis. Access to such information could facilitate the automatic assessment of quality in OERs and make them more easily findable. The authors of this work believe that availability of such information for educational resources will help make them more reusable. The authors conclude their observations by saying that "in order for information to be truly reusable for teaching and learning it needs to be annotated with descriptional and instructional metadata as well as content and domain metadata ". So, we could argue that the existence of metadata for an educational resource contributes to its reuse and thus to its quality as an OER.

In a recent study on the cost-efficiency and quality of OER integrated course materials[5], there is a definition for quality of learning material. According to this paper, **quality of educational material can be defined as appropriately meeting the stakeholders' objectives and needs which is the result of a transparent, participatory negotiation process within an organization**. The authors of this paper then stress that "some people are unable to believe that any process other than traditional peer review, licensing and publication can result in content that is highly accurate". From this sentence we can see that accuracy is one of the most important criteria in quality assessment. Furthermore, we agree that quality cannot be fully assessed without a human involved in the process due to the complexity of the concept and its dependence on the human perception. However we believe that some aspects of quality can be partially assessed without the help of human input. The results of this study showed that there were significant differences between the perception of teacher trainees and educators regarding the quality of educational material. This indicates that experience is a factor that affects how quality is perceived.

In 2006, a paper was published offering a new idea for a context-dependent ranking algorithm named LearnRank that focuses on learning applications[23]. The idea for this algorithm was inspired by the notion that quality is an intrinsically subjective concept as well as a characteristic of how a learning object is used in a particular context. The author of this paper starts by criticizing current strategies used to deal with quality issues claiming they are naive and misguided when used in large-scale repositories. We agree with this assertion since quality assessments are mainly dealt with manually (through peer reviews) and the amount of resources available in repositories is large and constantly expanding. The author begins by stating that based on her earlier experiences with ARIADNE (a learning object repository) she has started reconsidering the notion that quality assessments should be used to determine inclusion or exclusion of educational material from repositories. Furthermore she believes that quality assessments should not be dealt in a binary way (meaning to check if the quality criteria exist or not) but rather included as a factor that influences ranking of search results. The author then continues by claiming that current approaches in dealing with quality are unsuccessful because they avoid the real problem of measuring quality in context. She also claims that current processes are overly focused on less relevant quality aspects that are easily measured and processed. So, the author proposes a new idea "LearnRank", the idea that it is possible to automatically process real quality aspects. Next she provides a list of characteristics of quality, we present these in table 3.



Characteristic of quality	**Explanation**
Learning Goal	Depending on what a learner wants to learn about, **relevancy** is important. So, the learning goal helps separate the relevant from non-relevant material.
Learning Motivation	Depending on what the learner wants to achieve **(purpose)** certain materials are more appropriate than others
Learning setting	Depending on the **intention of use** certain types of material are more appropriate than others (e.g. provocative historical document not appropriate for unsupported self-study mode)
Time	Depending on how much time the learner has to achieve objective the **length** of the material is important. Does the learner want expansive, detailed material or introductory, short, concise material
Space	Connecting the learning experience to the physical environment can reinforce its effect. Ambient in which we learn affects quality perception (**NOT applicable for our purpose**)
Culture and language	Learning objects in native language of learner more easily assimilated than in other language they don't master. This also applies not only on geographically determined culture but also academic vs corporate culture, or learning context for engineers vs scientists etc.
Educational Level	Age and learning background affect the perception of quality
Accessibility	Educational material should be designed to cater to specific auditory, visual, motor or other needs. Designed for all

Table 3: Characteristics of quality and their explanations

The author of this work then advocates the need for **meaningful** ways to rank results so that we can deal with the vast amount of resources available. She proceeds by praising the PageRank algorithm used by Google, emphasizing how crucial the incoming links idea is in determining relevancy. Then she suggests the application of the idea of PageRank on learning objects (thus avoiding the need to ask for additional metadata), only instead of incoming links it should indicate how useful people have found a learning object. She believes that objects that are used in many contexts relevant to a specific learner should have higher LearnRank for that learner. Finally, she suggests the use of empirical data on the learning effect some objects like tools have in specific contexts in order to implement this idea in the future.

We believe that the basis of the idea for LearnRank has merit, since it mirrors the use of recommendation systems (which are already used for quality assessments) while narrowing down the context. However there are some weaknesses with this concept, firstly it does not help with the problem of discoverability. If a resource of the highest quality has not been discovered from the large amount of available educational materials and has not been used LearnRank will consider it irrelevant. Additionally, the fact a learning object is used by a large percentage of people in an institution in a specific context does not necessarily mean it is of highest quality. For example resource tools used in institutions are based primarily on financial considerations not quality ones. Also the idea of LearnRank focuses too much on context and not enough on the subjectivity of quality. Quality perceptions differ depending on an individual basis so what one person considers high quality might be of intermediate quality for another. Lastly, the implementation of a ranking algorithm like LearnRank presents a great challenge since it is difficult to account for all possible context scenarios.

In a recent state of the art review of quality issues related to OER[24], quality is described as an amorphous concept for which there is little experience and consensus on how to define and approach it. The authors of this review consider quality to be the "confluence of 5 concepts: *efficacy (fitness to purpose of object being assessed), impact (extent to which an object proves effective which depends on the object itself and the context in which it is used), availability (transparency and ease-of-access), accuracy (precision and absence of errors) and excellence (compare quality to its peers and its quality potential)*. In this work the process for assessing the quality of the content of courses is presented. Specifically a list of key features that are assessed automatically is mentioned, this list includes *metadata quality, grammar and language, tag quality, learning activities and technical correctness*. There is also mention of the various tools used for quality assessment, namely ratings, recommendation systems , social ranking, peer reviews and crowd-sourced peer-reviews. The authors also express their belief that when made available through process-controlled settings, OER can be assessed the same as traditional educational material (meaning compliance with set of formal quality standards and peer-review inspections). They also offer a list of quality criteria to assess learning objects, these are: *relevance to aims, cultural appropriateness, to communicate correctly within cultural context, type of institution, accuracy, importance, pedagogical effectiveness, completeness of available documentation*. As for OERs there are additional quality indicators for the effectiveness in the degree of openness, these are social, technical, financial and licensing. The paper concludes with some observations. The first that there is need for metadata that encompasses the concept of reusability in order to improve transparency and accessibility of OERs and facilitate their collection and resharing. The second observation is that the discoverability of resources hinges on the quality of their metadata. Finally the authors express their doubts about the adequateness of current quality assessment procedures for open education modules.

In the last two decades there have been two papers concerning the automatic assessment of the quality of educational materials using supervised machine learning methods. The first paper[25] concerns the automatic assessment of quality in order to test if a resource is sufficiently good to be used in a formal classroom setting for secondary science education. The goal of this paper is not the calculation of an overall quality assessment but the creation of a multi-dimensional characterization of the different aspects of quality. In this paper there is also mention of a list of quality assessment criteria identified from other sources, these are *accuracy, currency, relevancy, trustworthiness, readability, good general set-up, lack of bias and age appropriateness*. The first part of this work involved a study to identify the quality indicators that were most helpful in the acceptance or rejection of educational resources. The study identified the following indicators: *prestigious sponsor, age appropriateness, has sponsor, identifies learning goal, has instructions, identifies age range, organized for learning goals*. The results of this work showed that quality could be decomposed into meaningful dimensions, "more concrete" than the abstract concept of quality. Looking at the bulk of this work we agree with the decomposition of quality into different aspects for better assessment, however we believe these individual assessments could then be combined for an approximation of a more well-rounded quality assessment. Finally, we disagree with the assumption made in this work that quality is a binary attribute that either exists or not. Quality is a complex concept with degrees of variance as evidenced by the service quality model that compares perceptions with a preconceived notion of excellence.

The second paper[26] concerns the automatic assessment of quality in digital repositories in order to classify resources in different quality bands. In this work the authors propose an algorithm that tries to approximate the complex, time-consuming human judgements made when evaluating educational materials manually (these judgements are performed based on criteria like *relevancy, credibility, goodness, accuracy, currency, usefulness and importance*). To achieve this they first identify the various concerns regarding quality, namely issues about *accuracy of content, appropriateness to intended audience, effective design, info presentation, completeness of associated documents and metadata descriptions*. Then after a meta-analysis they used the following 16 criteria for the automatic assessment of quality: *cognitive authority (authority of the people that created the content), site domain (the domain shows if the resources come from educational and governmental entities), element count (number of elements in a resource's metadata XML file), description length, metadata* *currency, resource currency, advertising (number of advertisements on first page of resource), alignment (determine through metadata if resource supports national educational teaching standards), word count, image count, multimedia, www (social authority of a web page as measured by Google's PageRank), annotations, cost and functionality (the number of broken links and images that do not display on the first page of a resource).* Even though we agree that quality has variances, looking at the list of quality indicators we can see that the purpose and context of use were not factored in the proposed algorithm. Furthermore, we believe that in order to have an algorithm that deals with a variety of educational goals and contexts use of supervised machine learning algorithms is not a good idea because of the variety of input that would be introduced as training data.

2.4. Discussion

In this section, we summarize our findings from the literature review. First we provide some general observations on the topic of quality and how it is dealt with. Then we synthesize all the observations we made regarding quality and present a list of the factors that influence its assessment.

After assimilating all the information we get from the literature survey, we reach some conclusions. Firstly, there is a gap in the literature regarding the quality of educational material and the little literature that does exist is not used to its full potential. Our research on the the topic of quality has led to a common observation, **quality is a complex, multi-dimensional concept that is difficult to define and measure**. Assessments of quality for educational material mostly depend on manual reviews of the resources either by a recognized body or by users through systems equipped with rating and recommendation functionalities. The two attempts made to automate the quality assessment process,[25, 26] have some weaknesses. The first attempt treats quality as a binary value instead of a complex concept with variance. The fact we compare and make distinctions between products and services based on quality proves that it is not something that either exists or not. The second attempt completely disregards one of the main influencing factors of quality, namely the purpose for which a resource is intended. Additionally, due to its multi-faceted nature quality cannot be defined in an abstracted way but only in reference to a specific context. We firmly believe that quality should only be defined when the context is specified and after careful consideration of all the parameters (context, purpose, target audience and their needs and expectations). The literature suggests that due to the complexity of quality the majority of the attempts to achieve it resulted in a break-down of it aspects. This is evidenced by the variety of models and definitions[8, 11, 15] that focus on certain aspects of quality. However that encouraged the use of these models with a "use as needed" mentality that causes issues because it does not deal with the concept in a thorough, well-rounded way.

The majority of literature we found focuses on ways to assure, improve, measure and even automate quality, however there is very little research on understanding it. This results in a variety of work with scattered observations and suggestions regarding its meaning and how to achieve it. We continue by synthesizing all these observations and suggestions to list the various facets that influence and make up quality (thus answering research question 1a).

Facets of quality

1. *Quality is relative:* Quality is intrinsically linked with the satisfaction of the user meaning it depends on a person's perception of what constitutes quality [16, 20]. User satisfaction is an important factor in determining quality however it should not be the only one.

2. *Quality is context-dependent:* Besides relative quality is also context-dependent [15, 18]. In other words, based on who the interested parties are, their strategies, interests and expectations the indicators of quality differ. Quality cannot be defined out of context.

3. *Quality is purpose-dependent:* Similarly, quality depends on the purpose of use [16, 18]. Quality cannot be separated from the objectives of the users, therefore quality depends on how well the product or service (whose quality we are considering) helps the users achieve their goals.

4. *Quality is culture-dependent:* How a person views quality is subjective. Therefore perceptions are coloured by a user's upbringing, ideas and culture. Culture also refers to the cultural setting in which a product or service is used e.g. institution or industry [8, 23].

5. *Quality is dependent on education:* Since quality is dependent on the perception of the users it is logical that educational background also affects how they view quality[23]. Furthermore, the purpose for

which a product or service is needed can also determine the desired educational level of a product's or service's intended audience. For example if a professor is looking for material to create a lecture presentation for bachelor students, he/she is looking for material that were created having bachelor level education in mind.

6. *Quality is time-dependent:* The fact that people's perceptions change based on the experiences and knowledge they accumulate over time, means quality is affected by this factor. In addition, the context and goals change from one moment to the next which also indicates time is an influencing factor of quality.

7. *Quality is dynamic:* Another facet of quality is its need for continuous improvement and development, quality is not a fixed concept but rather a dynamic one [8, 15, 16]. Quality depends on a variety of factors like people's perceptions, objectives etc. These factors are not static, on the contrary they change with the passing of time and with them the users' view of quality.

2.5. CONCLUSIONS

In this section we use the insights and information of our literature review in order to answer the first research question of this thesis and more specifically sub-questions b, c and d. First we provide our definition of what quality of educational material is (answer to research question 1b) and then we answer the questions of whether quality can be quantified, to what degree (research question 1c) and which criteria can be used to assess it (research question 1d).

WHAT IS THE DEFINITION OF QUALITY OF EDUCATIONAL RESOURCES?

Before offering our definition of quality of educational resources, we first take a critical look of the existing definition (see section 2.3.3) to see if it satisfies our needs. This definition combines the essential ideas of three of the definitions presented in table 2, namely "Quality as the conformance to standards", "The good enough practice" and "Quality as meeting customer's stated needs". Although this definition is a fair approximation of what quality of educational material is it fails to fully capture the intricate nature of quality. More precisely, this definition is based on the assumption there is a static set of easily measured standards that can be used as a check list to determine whether the materials pass a certain threshold of acceptance. Therefore, this definition fails to encapsulate all the facets of quality.

Taking into consideration the findings of our literature review and the facets of quality we have identified we offer our own definition for quality of educational material:

Quality of educational material is the dynamic process that indicates whether this material satisfies an ever-changing list of requirements at a specific point in time. The existence of quality depends on people' perceptions that this resource is relevant and best suits their needs and objectives. Therefore as time passes and a person accumulates experience and knowledge and depending on the current trends and technological advancements his/her perception of what best satisfies his/her needs in accomplishing a goal varies.

IS QUALITY OF EDUCATIONAL RESOURCES QUANTIFIABLE AND TO WHAT DEGREE?

From our research on the topic of quality of educational material we have concluded that it is a mutli-faceted concept that is assessed by a variety of parameters. This is evidenced by the numerous factors that influence its definition as well as the various lists of quality assessment criteria presented in the literature. In addition, we have ascertained that it should not be treated as a binary value. On the contrary, any attempts to quantify quality should result either in a number on a scale or an ordered category. We come to this conclusion from some service quality models that depend on comparison presented in section 2.3.2, the classification of resources in quality bands on the paper about automatically assessing quality[26] (see section 2.3.3) and the stated need for easy ways to find and distinguish high quality materials.

In order to answer whether it is possible to quantify quality we observe the findings of our literature and more precisely the indicators of quality that we identified for quality assessments. These quality indicators (criteria) are the following:

QUALITY CRITERIA:

1. *Relevance:* One of the most important factors in determining the quality of learning materials is their relevancy to the topic the user is interested in. Since quality is purpose-dependent it is obvious that resources that are irrelevant to a user's search query will automatically be considered of low quality. The

importance of relevancy is also evident from the fact that search engines rely on this factor to provide search results. Relevance can be separated in topical and user relevance (see section 1.2). **Topical relevance** is of primary importance in regards to educational resources because they indicate the field of study the resources were created for. Its importance is also evident from the fact some retrieval models focus entirely on it.

2. *Content (Well-Organized, Comprehensible and Well-presented):* One of the most important criteria for deciding whether a learning resource is of high quality is its content. The importance of this criterion is evidenced by the re-occurring mention of parameters regarding content in the lists of criteria we found from literature. More precisely research suggests that a high quality resource should have a well-organized, logical structure and comprehensible, well-written content. In addition, it should be presented in a pleasing manner. Unfortunately these aspects of a learning resource are difficult to determine due to their subjectivity and cannot be translated directly into a value. However considering the importance and multitude of parameters covered with this criterion (structure, design, comprehensibility, well-written/well-presented etc.) we can use available reviews, ratings or recommendations to proxy its quantification.

3. *Reliable:* Another important factor that shows if an educational material is of high quality is its reliability, meaning how accurate and trustworthy the information it contains is. This is another criterion that is extremely difficult to pinpoint and measure. Therefore to measure it one can rely on an approximation by using either the author's, institution's or platform's reputation (the same way some of the papers on assessing quality have done previously). After all, highly recognized institutions/platforms and people with expertise on a topic should provide the most accurate and reliable content.

4. *Up-to date:* The date of when a resource was created or last updated can be used as a quality measure since current and up to date materials are more useful.

5. *Length:* The duration or length of an educational resource can also be used as an indicator of how detailed the given information is. This will allow the division of the materials to those that only offer essential explanations on the subject matter from those that give a more detailed, expansive one.

6. *Number of topics:* A resource could contain information regarding several subject matters, not just one. Therefore this criterion could be used as an indication of whether a resource is useful for introductory or more in depth learning purposes. This will allow us to better cater to the purpose of use which we identified as one of the facets of quality.

7. *Culture and Language:* Quality is dependent on the culture and language of the users. People with different cultural backgrounds have very different standards regarding what constitutes quality. This is true not only for geographically determined differences in culture, but also for differences between other kinds of cultures, such as the academic versus the corporate culture etc. In addition, language is another indicator of quality since people are predisposed to prefer material in their native language rather than in a secondary language they may not master. It is important to note that this is not the case for all fields of study, for example due to their constant interaction with computers, computer scientists prefer material written in English.

8. *Educational Level:* Depending on their educational level users judge quality differently. For example material that is appropriate and considered of high quality for a high school student is considered of low quality and of low use to a Phd student that has advanced knowledge on the topic. Therefore depending on the learning goal, if we can also identify the target audience or the objective for which it was created we are provided with another important indication of quality.

9. *Copyright Licenses:* Similarly another indicator of high quality for OERs are their copyright licenses and more specifically the type of Creative Common licenses used (see appendix A). When trying to reuse, modify or redistribute freely available resources restrictions are imposed on the actions you can take by the CC license associated with them. Hence a CC0 license which allows unrestricted use of the material indicates the highest quality whereas a C license that reserves all rights indicates lowest quality.

10. *Editable:* By their very definition open educational resources subscribe to David Wiley's framework of openness known as the five Rs. According to this framework open resources need to be freely accessed

and shared so they can be reused, revised, remixed, redistributed and retained. Therefore another indicator of high quality is whether a resource can be edited so that it can be modified and reused. For example educational material with logos or digital watermarks splashed across them are less useful and of lower quality when the user needs to extract part of the content.

11. *Technical characteristics:* With the advancement of technology, educational materials are no longer restricted solely on the written word but rely on audiovisual aids as well. For these types of materials an important factor that determines their quality is the image resolution for pictures, audio quality for recordings and a combination of the two for videos. It stands to reason that images, recordings and videos of poor quality that make it hard for the user to discern the content are not very useful neither for teaching nor for learning.

12. *Existence of images:* In the literature it is mentioned that we should cater to the needs of our target audience. Depending on the person, the way he or she assimilates information varies. For most people the existence of visual aids is important in understanding. So the existence of images is another criterion that influences the quality of a resource.

13. *Accreditation:* If a resource has been accredited by a recognized body this provides an indication of high quality. Furthermore if the accreditation is accompanied by annotations made from the peer review team that would provide further proof of the quality of the material.

Looking at these quality criteria we see that there are certain criteria that can be quantified (e.g. length, up-to-dateness) and other that cannot (e.g how well-organized, comprehensible or reliable a material is). We can easily compare the dates of resources to distinguish which ones are more current, we cannot however quantify or compare the organization/structure or comprehensibility of resources. Therefore we conclude that it is impossible to fully convert quality into a number that will fully grasp its definition. However we can partially achieve quantification of quality by using the easily quantifiable quality criteria and by using approximations for the more complex, intricate quality criteria.

3

From Theory to Practice

In this chapter we present the process we followed to translate the findings of our literature survey into practice so we could automatically assess the quality of educational material and reorder them based on it.

3.1. Introduction

After conducting the literature survey presented in chapter 2, we were able to answer the first research question of this project. Namely, we identified the factors that influence the perceptions of quality, we provided our own definition of what constitutes quality of educational resources and produced a list of criteria that could be used to partially quantify it. In order to answer the remaining questions of this project we need to find a way to automatically assess the quality of OERs and then use this assessment to create a more meaningful way of re-ranking them. To achieve this we first have to determine the various facets of quality for our goal. Then we need to find a way to get the information we need regarding the quality criteria for each resource. Having done that we need to determine how to use this information in an automated way to produce a quantification of quality (thus answering the second research question of this thesis). After that we need to decide how the re-ranking should work and create an algorithm to achieve it. Finally, experiments are needed to test whether the automatic quality assessment and consequently the re-ranking proposed are successful. For the assessment of the re-ranking we obviously require a search engine on which we can integrate this process (meaning our quality model for automatic assessment of quality and re-ranking). Fortunately, we were able to gain access to a prototype search engine created by the company Feedback Fruits.

3.2. Determine facets of quality

In this section we determine the various facets of quality for our goal. From the literature survey we concluded that to properly define and achieve quality all the influencing factors should be carefully considered in relation to the context. So, here we determine the various facets of quality in order to correctly assess it for open educational resources.

- **Quality is context-dependent:** In order to properly define and assess quality, first the context should be determined. In this project, our context is the following: Help educators who are searching for high quality material for a variety of educational purposes find them more easily.

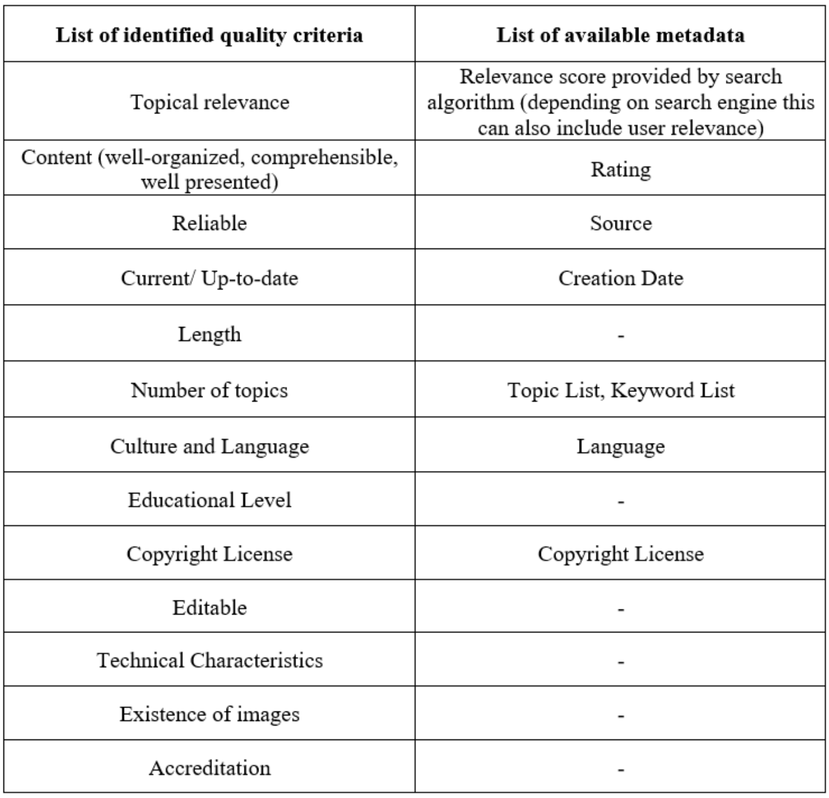
- **Quality is relative:** Quality is intrinsically linked to the satisfaction of the person assessing it. So, we need to determine who our target audience is and then gain an understanding of their needs, thoughts and expectations. For the purposes of this thesis we focus on 3 groups of people: professors, teaching assistants and Phd students. These are groups of people that are likely to search educational material for a variety of reasons and require help in finding high quality material. In order to gain insights on the needs, expectations and processes our target audience follows when searching for educational material we conducted a user group analysis and gathered the needed information (see the following chapter, chapter 4).

- **Quality is purpose-dependent:** Perception of quality also varies based on the purpose of the users. Therefore we need to narrow down what the resources searched for by our users are needed for. For the purposes of this thesis we will only consider two educational goals that require search of high quality material, **Prepare a lecture presentation** and **Expanding knowledge**.

- **Quality is culture-dependent:** Culture also plays a factor in how quality is perceived. For our purposes we use the cultural context set by our institution, TU Delft. Specifically, according to TU Delft's policies and culture the use of the English language in most settings is required, to accommodate its multitude of international students and personnel. So, for our project we limit the search for resources to only English material.
- **Quality is dependent on education:** In order to successfully cater to the needs of our target audience we need to consider their educational level as well as the educational level the resources cater to. As far as the educational level of our target users is concerned by narrowing down the group of users we have already determined their educational level. So, we cater to this criterion with the user group analysis we conducted (see chapter 4). As for the educational level the resources cater to that is why educational level is included in the list of quality criteria used to assess the quality of resources.
- **Quality is time-dependent and Quality is dynamic:** These two facets are intrinsically linked so we cater to them together. Specifically, since quality perceptions are constantly changing and depend on time the formula used to assess quality should be flexible and easily adaptable. In addition, this formula to assess quality should be changeable and cater to quality calculations on a case by case basis.

3.3. META-ANALYSIS

In this section, we address the issue of getting access to the information we require for each resource. In order to get this information we propose the use of available metadata attached to resources. For this reason we examined the metadata that are provided for educational resources from three open educational repositories, MIT OpenCourseWare, TU Delft and OER Commons. Then we compared the list of available metadata with the list of criteria we identified from the literature survey to see which ones are available and can be used for automatic assessment of quality (answering research question 2a). The comparison between the list of identified criteria and list of available metadata is shown in table 4.

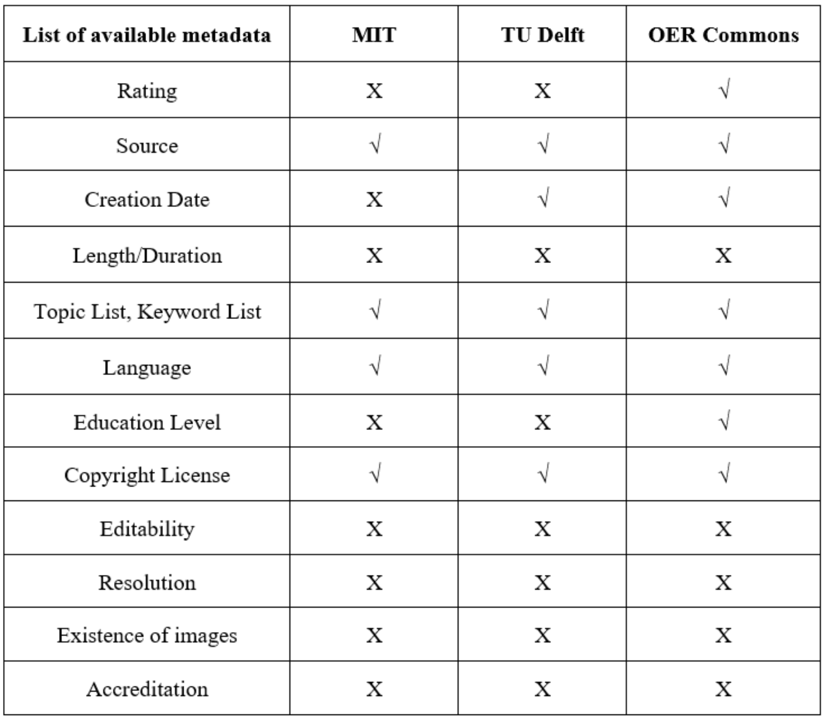


List of identified quality criteria	List of available metadata
Topical relevance	Relevance score provided by search algorithm (depending on search engine this can also include user relevance)
Content (well-organized, comprehensible, well presented)	Rating
Reliable	Source
Current/ Up-to-date	Creation Date
Length	-
Number of topics	Topic List, Keyword List
Culture and Language	Language
Educational Level	-
Copyright License	Copyright License
Editable	-
Technical Characteristics	-
Existence of images	-
Accreditation	-

Table 4: Comparison of list of identified quality criteria against available metadata

3.4. RESOURCE PROCESSING

When going through the list of available metadata from these three repositories we observed that the list provided is very limited and the existence of metadata sparse. In addition, the list of available metadata presented on table 4 reflects the metadata available in all three repositories. However individually some of them were missing important metadata tags. The list of available metadata in each repository is presented in table 5.



List of available metadata	MIT	TU Delft	OER Commons
Rating	X	X	√
Source	√	√	√
Creation Date	X	√	√
Length/Duration	X	X	X
Topic List, Keyword List	√	√	√
Language	√	√	√
Education Level	X	X	√
Copyright License	√	√	√
Editability	X	X	X
Resolution	X	X	X
Existence of images	X	X	X
Accreditation	X	X	X

Table 5: List of available metadata per repository.

From table 5, it is evident that the OER Commons repository has the most well-maintained metadata collection. Unfortunately, due to a miscommunication with OER Commons we were given access to their metadata too late in the project which resulted in them not being included in our implementation. This also caused the quantification formula we use, to include the rating criterion even though the information was not available for any of the resources we had access to. So, after identifying this lack of enough available metadata from MIT and TU Delft as well as the sparsity of the metadata that are provided, we decided to process the resources offered by these repositories and extract any missing information that we could (unfortunately that only includes length and creation date). So, for the TU Delft repository we processed 58729 resources in order to get information about the length criterion. Specifically, we calculated the number of pages of each document. Unfortunately, even after our processing not all resources had a complete list of the needed information available (some resources were missing keywords list, creation date etc.).

As we mention above in order to test the quality model we propose, we needed to do some experiments. Therefore it was necessary to implement a prototype search tool that automatically assesses the quality of open educational resources and reorders them accordingly. In order, to accomplish this a prototype search engine was made available to us from the company Feedback Fruits. Specifically, a GraphQL access point to their search engine was made available to us and from it we also gained access to resources from the MIT OpenCourseWare repository. Therefore, we only have access to the metadata of these resources after performing a query. So pre-processing the needed information like we did for the TU Delft resources was not possible. Instead we perform an on the spot processing of the resources to get information regarding their creation-date. Specifically, the creation date of most MIT resources is included in their url so we process it to get the information we need.

Having identified the list of available criteria that can be used for automatic quality assessment, we then needed to see which combination of these criteria makes up our target audience's quality perception. This is another reason for which we needed to do the user group analysis presented in the following chapter (see chapter 4). Because we needed the information and insights gathered from it to design and implement our quality model(see chapter 5).

4

User Group Analysis

One of our findings from the literature survey in chapter 2 is that quality can only be defined in a specific context after careful consideration of all the facets of quality. Therefore to ascertain how quality is perceived by our target audience and better design a quality model that automatically quantifies the quality of educational material, a good understanding of the users is needed. For that reason a user group analysis was conducted. Specifically we had eight interviews with some members of the personnel of TU Delft. From these interviews we were able to corroborate some of the findings of our literature review and gain useful information for the design of the quality model we present in a following chapter (see chapter 5). Below we describe in more detail how these interviews were conducted and the results we got from them.

4.1. Purpose of Analysis

In order to design an effective quality model that automatically assesses the quality of educational material a better understanding of the target audience is needed. Specifically, we need to gain insight on the needs and expectations of our users (meaning professors, teaching assistants and Phd students). We also need information on the processes they follow when looking for educational material. Additionally we want to corroborate some of the findings of our literature survey since the literature focusing on the quality of educational material was so limited. Finally, we want to ascertain how our target audience perceives quality. For this reason they were asked to place in order of importance the list of quality criteria we identified from our literature review (for this list we used the criteria for which we had or thought we had available metadata). Specifically, we want to see which criteria are considered more important by our target audience which also gives as an idea on how they perceive quality (since the criteria essentially make up quality).

4.2. Hypothesis

We believe that the interviewees will corroborate the findings of our literature survey in regards to the definition of quality of educational material and confirm the difficulties they face in finding high quality educational material. We also hypothesize that the ordering of the quality criteria from the participants will be different from person to person due to the varying perceptions of quality. However we believe that rating and source will prove to be the most important criteria from the orderings since these are the quality criteria we encountered more often in the literature.

4.3. Measurements

From these interviews we plan to get a variety of definitions on quality of educational material from the perspective of our target users. We also want to gather information on the process they follow when looking for educational material for two separate educational purposes. Furthermore we want to obtain information regarding their needs and expectations when looking for material for these distinct educational purposes. Finally we want to make a qualitative assessment of the orderings of the quality criteria in order to get a weight of importance for each criterion.

4.4. Set up

4.4.1. Participants

These interviews were conducted with the help of some personnel members from TU Delft. Specifically, the participants of these interviews were one Teaching Assistant (T.A), three Phd students, three Assistant Professors and one Associate Professor. The majority of the interviewees have a background in Computer Science and Engineering, one has a background in Aerospace Engineering and another in Marine Technology. The participants all came from differing cultural backgrounds. Some were Dutch, Greek, German, Canadian, Chinese, Iranian and Mexican.

4.4.2. Set up

Before conducting the interviews with our participants we created a questionnaire (see appendix B) with all the information we wanted to obtain. These questions were separated in 5 sections. The first section concerned the collection of some general information. Specifically, the first question was designed to gain insight on the process the participants follow to select their material. The second question was designed to ascertain if the educators considered that finding high quality educational material is an issue. The next section contains only one question to corroborate the findings of our literature survey. This is designed to see what the interviewees consider quality and if it is in line with our definition. The following section contains two questions to check the importance of the criteria copyright license and language. These questions were added because these are quality criteria that were identified by only one source of literature. Next there is a question about ordering the list of quality criteria in order of importance when assessing quality of an educational resource. This question is designed to provide us an idea of how the target audience perceives quality. Next follow two sections, one focused on the learning purpose "Lecture Presentation" and the other on the learning purpose "Expand Knowledge". Both of these sections contain the same questions designed to gather information on the expectations, needs and processes of the interviewees when dealing with these tasks. The section for "Lecture Presentation" contains an additional question to check if there is a preference for multimedia educational content for this task.

4.5. How was the analysis conducted?

The interviews were conducted one on one in a structured manner following the questionnaire (see appendix B) described in the previous section. The interviewees were asked to answer the questions in the same order and their replies were written down. For the task of ordering the quality criteria in order of importance, the interviewees were supplied with cards with the criteria written on them and asked to place them in order of importance. Although the interviews were structured, the participants were encouraged to elaborate on their habits, thoughts and processes. This way we could gain a better understanding of them, their needs and their expectations.

4.6. Results

In this section we present the results we got from this user group analysis. Specifically, we concentrate on the corroboration of our literature's findings, on any insights we get regarding our users' needs and processes and lastly we present the qualitative assessment of the orderings provided by the participants in order to assign weights of importance to each quality criterion.

Corroboration of literature findings

First, we focus on the answers we got for the questions that were asked in order to corroborate the findings of our literature review. First we start with the question on defining quality of educational material, then we continue with the question regarding copyright licenses and lastly the question asked regarding culture and language.

When asked to define quality of educational material, **seven out of eight** participants started naming a list of requirements they wanted in a resource. The criteria mentioned by these participants are: *source, size, creation/modification date,clarity of written word, well-written, well-organized, contains a lot of figures, graphs and images, clarity, right level of education, clearly explained, well-structured, looks nice, contains well-designed experiment, concise, good resolution, easy to understand, effective, good technical characteristics, good content.* Only one of the participants provided a more abstract definition of quality but even in this definition specific requirements were mentioned. More specifically, this participant defined quality of educational material as *truthful statements that help their audience to become more inquisitive, innovative and develop their masters skills.* This participant also listed three additional requirements, namely that high quality material should be *reliable, approved through peer review and of the right educational level.* Although the list of requirements varied from person to person there were some characteristics mentioned by all of the remaining 7 participants. Namely, *well-written and comprehensible content.*

Copyright license is a criterion we only identified once as part of a list of quality assessment criteria in the literature. However, the definition of open educational resources and the proponents of open education seem to place great importance on it. Therefore we asked the participants whether copyright license was one of the deciding factors for them when selecting educational material. **Two out of eight participants** responded that they consider copyright license to be of great importance, one of them because of his involvement in MOOCs and the other because of the restrictions imposed by the copyright licenses. **Six out of eight participants** responded that they did not consider copyright license as important although three of them added a caveat. One of those three said that copyright license becomes a factor depending on the purpose of use of the educational materials e.g. for the preparation of a public lecture it is a factor. The other two participants said they tend to create their own material for lecture presentations so copyright license is not an issue in that regard, however when looking for freely available papers for other purposes it is.

We also asked the participants whether they prefer educational material in their native language when trying to expand their knowledge. All of the participants responded that they prefer material in English. However one of them continued by saying that there are not good enough material in his native language otherwise he would prefer it. For one of the participants English is his native language.

INSIGHTS ON TARGET AUDIENCE

In this section we present the answers we got from the participants that were designed to give us insight into their needs, processes and expectations when looking for educational material.

When asked whether finding high quality material is a difficult process, **six out of eight** participants responded in the affirmative. From the remaining two participants, **one** answered negatively and **the other** indicated that in his field of expertise he had no difficulties because over the years he had established a network of peers that collaborated, however for other fields he found it difficult. **From the six** interviewees that responded in the affirmative **two** specified that they found the process of finding high quality material very time-consuming. Also **three out of those six** participants stressed that it was difficult to find quality material for more advanced and recently developed topics. **Two out of these six** participants speculated that the reason for this difficulty in finding high quality resources originates from legal restrictions imposed by organizations and universities regarding access to their resources.

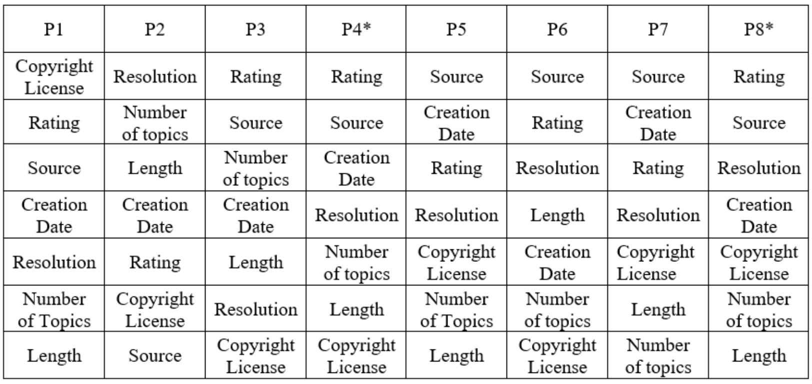
The participants were also asked to elaborate on their habits for determining which resources were useful or not when searching for educational material. Specifically, the participants were asked whether they read the educational material they find online or download it first. **Four out of eight** interviewees said they usually depended on the description or abstract of the resource to determine its usefulness. From the remaining four, **three** said they usually look at all the content of a resource on the spot to decide. And the other participant said he either looks at the whole content or only parts of it.

During the interviews the participants were presented with two educational goals that require looking for educational material. Then based on these they were asked the type of resources they look for to achieve each goal. The two educational goals presented to the participants are: a) Preparing a lecture presentation and b) Expanding their knowledge. **Five out of eight** interviewees said that for the preparation of a lecture presentation they would look at existing lecture slides, videos and images. The other **three** said they would look for textbooks and papers. For the purpose of expanding knowledge all participants said they would look at papers, textbooks and conference proceedings. **One** of them also mentioned lectures while another mentioned anecdotal stories from his colleagues.

The participants were also asked how they would proceed if they could not find the material they were looking for. **Four out of eight** participants indicated they would settle for whatever was available especially if they had other pressing work. **The other four** indicated that rather than settle and adjust they would create their own material. Also **six out of eight** participants said they were willing to freely share any material they created.

QUALITATIVE ASSESSMENT OF ORDERINGS

In this section we present the results we got from the participants regarding the order of importance they would place the list of quality criteria. This list is comprised of the metadata we identified as available at



P1	P2	P3	P4*	P5	P6	P7	P8*
Copyright License	Resolution	Rating	Rating	Source	Source	Source	Rating
Rating	Number of topics	Source	Source	Creation Date	Rating	Creation Date	Source
Source	Length	Number of topics	Creation Date	Rating	Resolution	Rating	Resolution
Creation Date	Creation Date	Creation Date	Resolution	Resolution	Length	Resolution	Creation Date
Resolution	Rating	Length	Number of topics	Copyright License	Creation Date	Copyright License	Copyright License
Number of Topics	Copyright License	Resolution	Length	Number of Topics	Number of topics	Length	Number of topics
Length	Source	Copyright License	Copyright License	Length	Copyright License	Number of topics	Length

Table 6: Ordering of quality criteria in order of importance (the asterisk * shows, orderings for which the participant considered the last three criteria of equal importance)

the time for automatic quality assessment. This task was asked in order for us to ascertain how our target audience perceives the quality of educational resources. In table 6 we present the results we obtained from this question.

4.7. INTERPRETATION OF RESULTS

CORROBORATION OF LITERATURE FINDINGS

In this section, we analyse the results we got from the questions that concerned the corroboration of the findings of our literature review.

From the question where participants were asked to define quality of educational material we get confirmation of our definition. Specifically, the fact that all 8 participants mentioned varying lists of requirements that characterized their view of high quality, is evidence we are on the right track with the definition we propose. In addition, almost all the requirements mentioned by the participants are criteria we identified through literature. The only criterion we overlooked was the "well-designed experiment". In the literature there was mention of need for scientific foundation however we interpreted that as reliability of the content instead of a separate requirement.

Going over the answers given on the topic of copyright licenses and whether its a deciding factor in the selection of educational material we can see that another finding from our survey is corroborated. Namely, the fact that according to the purpose of use the definition of quality for educational material changes. From the answers we collected it is clear that copyright license increases in importance when the material needs to be reused.

The expressed preference for material in the English language is in line with our observations regarding the culture of TU Delft since English is the preferred language for communication. However, the responses for this question were biased because the majority of the participants work in the field of Computer Science, in which English is preferred anyway.

INSIGHTS ON TARGET AUDIENCE

In this section we analyse the responses we got regarding the habits and processes the participants follow when looking for educational material for specific purposes. Here we try to glean any information that might prove useful in the design of our quality model.

Based on the answers we got regarding the difficulties in finding high quality educational material, we observe that it is considered a difficult, time-consuming process. So, there is a need for the goal we are addressing with this thesis, namely there is a need for more meaningful ways of ranking search results.

From the responses gathered regarding the habits of participants when looking for educational material in general, we conclude that the majority of our target audience rely on the abstract or description of a resource to quickly sift through them. This is an insight that proved useful in our implementation.

Regarding the types of materials participants prefer, we can observe that there is a clear distinction between the type of educational material that is sought depending on the purpose of search. More precisely, it is apparent that the majority of our target audience look for existing lectures, videos and images when searching for educational material to prepare a lecture presentation. However for the purpose of expanding knowledge the preferred type of educational materials are papers,textbooks and conference proceedings.

Another useful insight in our target audience's habits is the way they respond to the absence of high quality educational material. Specifically, half of them settle for whatever suits their needs from existing materials while the other half create their own. Regarding the first half of the participants this is an expected behaviour considering the service quality model regarding benefits and sacrifices described in section 2.3.2. Based on this view it is clear that filtering out educational resources based on quality judgements is not a good idea. The willingness of the majority of the participants to freely share their materials could prove very beneficial for open educational repositories in the future. More specifically, if educators share freely their own high quality materials this could improve the impression some people have that there are not many high quality OERs.

QUALITATIVE ASSESSMENT OF ORDERINGS

Here we look at the 8 different orderings provided by the participants and qualitatively assess how they perceive quality of educational material. More specifically, based on these orderings we assign weights of importance to each criterion. It was an oversight on our part not to ask our participants to offer two separate orderings of the results one for each of the presented purposes. However from their answers regarding copyright licenses we were able to conclude that for the purpose "expansion of knowledge" the weight assigned to this criterion should be smaller than the one assigned for the preparation of a lecture.

Based on the responses presented in table 6 we can see that the criteria source and rating are considered the most important ones. This is evidenced by the fact they are both listed first in three out of the eight orderings and were placed in the first three places in seven out of the eight orderings. Next the criteria creation date and resolution seem to be the ones placed in higher order since they are placed in the first four places six out of eight times. Then the criterion regarding copyright license appears to be perceived as slightly more important since one person even placed it in the first place. In addition, since we are trying to automatically asses the quality of Open Educational Resources copyright license is an important factor, so we place it next. Finally with so few orderings we cannot make a clear distinction between the remaining two criteria Number of Topics, and Length. So, our hypothesis that source and ratings would be the criteria placed higher in the orderings is confirmed. Considering all of this, the weights we assign to these criteria are: **Source=25%, Rating= 25%,Creation Date=15%, Resolution = 15%, Copyright License=10%, Number of topics = 5% and Length=5%**.

After we conducted these interviews the library of TU Delft, organized a panel regarding the topic of quality. During the presentation the attendees were asked to select which of a subset of these criteria they considered most important. The results of that panel were: Up-to date= 19%, Source = 11%, Length = 1%, Copyright License = 11% and Resolution=5%. Due to the fact we were only able to conduct a few interviews as part of the user group analysis and the panel's attendees were mainly educators of TU Delft we decided to use the above results and create a second set of weights. What we can glean from these results is that creation date, source and copyright license are considered the most important criteria while resolution is perceived as of little importance in comparison. So, combining these observations with the results of our user group analysis we offer a slightly different view of quality: **Source=20%, Rating= 20%,Creation Date=20%, Resolution = 10%, Copyright License=20%, Number of topics = 5% and Length=5%**

4.8. THREATS TO VALIDITY

In this section we present the threats to validity we are faced with in this user group analysis. Firstly, the number of reviews conducted were too few to give a representative view of our target audience. Furthermore, the majority of the participants have a background in Computer Science which could skew the results e.g. there was a bias in the answers regarding culture and language. Also the perception of quality we identified through qualitative assessment is weak, a larger sample of participants should have been used so a quantitative assessment could be performed. Finally, due to quality's dependency to the purpose of use and our observation regarding properly defining quality only in reference to a specific context, the participants should have been asked to offer two orderings with the two different purposes in mind each time. This oversight led to the assumption that our target audience's perception from one purpose to the other only changes in regards to the copyright license criterion.

4.9. Conclusions

Although the participants of this user group analysis were very few we were still able to get the information we required. Namely, we were able to get confirmation of our definition for quality of educational resources, of the dependency of quality to the purpose of use and of the quality criteria we identified through our literature survey. In addition, we gained insights that are useful for the design of our quality model. As implied in our literature review it is evident that the purpose of search influences perceptions of quality. Specifically, for the two purposes used in this analysis we found that the importance assigned to some criteria e.g copyright license changes. We also found that based on the purpose of search there are certain preferences for specific types of materials. We also gained insights that might prove helpful in the implementation of our prototype search tool. For example the reliance of the target audience in the abstract or description to quickly decide whether they consider a resource to be high quality or not is an interesting discovery. Finally, we were able to qualitatively assess the importance of criteria for our target audience. That way we were able to determine how our target audience perceives quality in order to use it in our quality assessment calculations.

5 Methodology

Observing the various facets of quality and having defined quality of educational material it is clear that automatically assessing its full scope is an impossible task. Quality is too abstract of a notion to quantify since it changes according to the context, purpose and person perceiving it. However it is possible to partially quantify quality and use this value as an indication of the existence of high quality in a resource. Therefore the next step is the conceptualization of a model that partially quantifies quality of educational resources and then reorders any available search results accordingly. In the following sections I present the conceptual design of a quality model that achieves this goal.

5.1. Prerequisites

Considering carefully the results of the literature survey on the topic of quality and the user group analysis, there are certain requirements that need to be met to build the aforementioned quality model for educational resources. Namely, the facets of quality have to be determined. This was done in chapter 3 section 3.2. From these facets it became clear it was necessary to obtain the target audience's perception of quality, this was achieved through the user group analysis (see chapter 4) where we identified two sets of weights for the available list of criteria.

Having defined the facets of quality and determined the perception of quality by our target audience in regards to certain educational purposes, we consider what else is needed to create a quality model that automatically assesses the quality of OERs and re-ranks them accordingly. The one thing missing is the inclusion of the most important quality criterion, relevance and more precisely topical relevance. Relevance (or just topical relevance depending on the search engine) however can only be calculated based on a search query, so we realize the only other thing needed is a search engine that provides an initial list of documents ranked based on their relevancy. Therefore any quality model we design can only exist if integrated on a search engine.

Considering the design of our quality model we therefore conclude its starting point is the acquirement of a list of relevant (or topically relevant) documents ranked by relevancy. We also need the relevancy score of each document in order to combine it with the partial quality score assessed (this partial quality score comes from the criteria we identified as available in chapter 3). That means that the remaining parameters we need to determine in our quality model is how to produce a partial quality score for each resource and how to combine this score with a resource's relevancy score to produce a new ranking of the resources (thus answering research question 2b). In the following sections we describe in detail the conceptual design of this quality model.

5.2. Diagram of Quality Model

Figure 5.1 shows the diagram of the quality model we designed and implemented in order to automatically assess the quality of resources obtained by a search query and re-rank them based on it.

This quality model is built in order to facilitate the workload of TU Delft educators who are looking for Open Educational Resources (OER) for a variety of educational goals. It is built taking into account the policies and culture of TU Delft. In particular, the fact that courses in TU Delft are all taught in English.

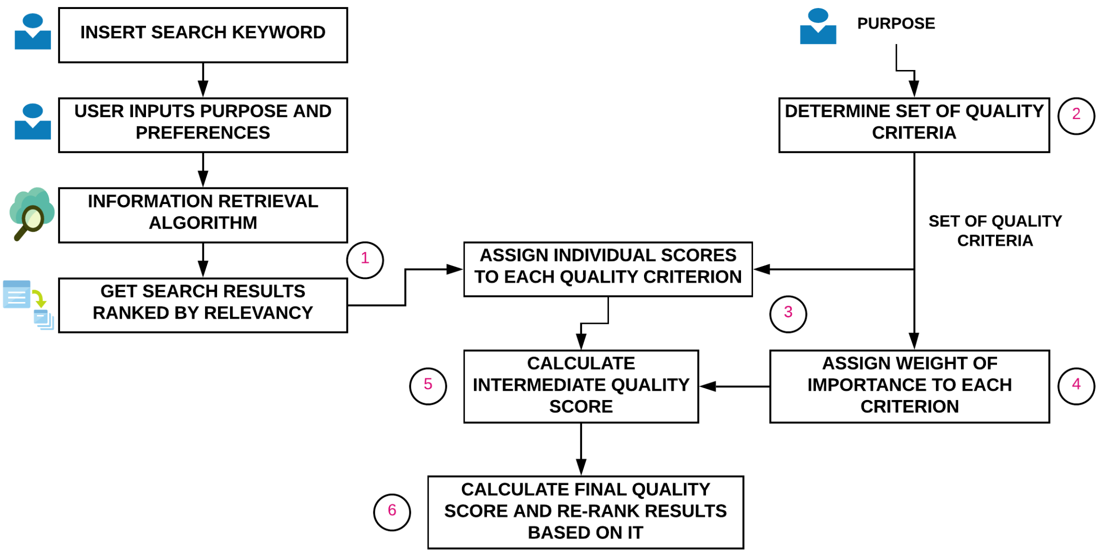

INSERT SEARCH KEYWORD
USER INPUTS PURPOSE AND PREFERENCES
INFORMATION RETRIEVAL ALGORITHM
GET SEARCH RESULTS RANKED BY RELEVANCY
1
PURPOSE
DETERMINE SET OF QUALITY CRITERIA
2
SET OF QUALITY CRITERIA
ASSIGN INDIVIDUAL SCORES TO EACH QUALITY CRITERION
3
ASSIGN WEIGHT OF IMPORTANCE TO EACH CRITERION
4
CALCULATE INTERMEDIATE QUALITY SCORE
5
CALCULATE FINAL QUALITY SCORE AND RE-RANK RESULTS BASED ON IT
6

Figure 5.1: Diagram of my quality model

5.3. Conceptual Design of Quality Model

In this section we present the conceptual design of our quality model, specifically we break down the quality model in six steps and explain them in more detail. For each step we present the design space, our design choices and offer possible alternatives.

5.3.1. Step 1

The first step of the model consists of the use of an existing search engine to retrieve all relevant results of a query. In addition the purpose of search and preferences of the user regarding quality (meaning the user's perception of quality) are inserted at this point. This is necessary because a quality model needs to cater to different people with varying perceptions on the topic of quality and with different purposes in mind.

Design Space for Purpose of Search:
There is a multitude of reasons why a person may search for educational material. So considering the context in which this model is designed we present the most common purposes for which educators may perform a search. These purposes are separated in two categories, each representing a similar outlook in the type of material the user is looking for.

1. Preparing for presentations, teaching and scientific writing
   - Preparing a lecture presentation
   - Preparing a conference presentation
   - Preparing for a class
   - Preparing an exam
   - Searching material to give background knowledge to students
   - Researching in order to write a scientific article
   - Researching in order to write a textbook
2. Seeking knowledge
   - Researching in order to expand knowledge on a topic
   - Researching in order to remember prior knowledge on a topic
   - Searching material to expand digital learning environment etc.

Design Choice:
When implementing this model we selected two of the aforementioned purposes, more precisely we selected the **Presenting a lecture presentation** and **Researching in order to expand knowledge on a topic** purposes. These purposes were selected as representative samples of the two categories of educational goals.

5.3.2. Step 2

After retrieving all the relevant results from a search query, the user input regarding the purpose of search is used to determine the set of criteria used to quantify quality. Based on the purpose of search some of the criteria of quality become less important or even obsolete. Additionally, the desired value for some of the criteria changes based on the purpose of search. Specifically, for the preparation of a lecture we require all the criteria whereas for expanding knowledge the criteria editable and copyright license are of lesser importance and could be removed from the list of criteria we use. Alternatively, a change in the weight assigned to them could achieve the same result.

Design space for Set of Criteria to quantify quality:
Based on which of the two categories the inserted purpose of search belongs to the set of criteria used to quantify quality changes. Specifically, if the purpose of search belongs to the second category **Seeking knowledge** then the criterion about the copyright license is of lesser importance and could be removed from the set of criteria to quantify quality. Alternatively, instead of removing this criterion from the set of quality criteria it could be given less importance when quantifying quality.

Design Choice:
For our implementation we decided to keep the same set of criteria for all purposes and adjust the importance attached to them accordingly. We believe that keeping the same set of criteria for all purposes will provide a more accurate value since quality is a multi-faceted concept and all components influence it. Furthermore, while searching for material people sometimes get sidetracked and change the purpose of their search so calculating all the quality related information is better.

5.3.3. Step 3

From the literature survey on quality a list of possible criteria for the quantification of quality is presented (see section 2.5). In this step each criterion is quantified and assigned an individual score. This score indicates if this criterion is close to a desired value, which in turn indicates high quality of a resource regarding a certain characteristic. At this point there are two challenges to be faced: a) how to assign these scores to each criterion and b) at which point should the quantification of each criterion occur. Below we present the design space and our design choices for each challenge.

Design Space:
a. How to assign an individual quality score to each criterion:
There is an infinite amount of values and ranges that could be used to assign an individual score to each criterion. Some possible values and ranges are:

- {0,1}
- {0,10}
- {0,100}
- {-100,100}

As for the assignment of a score this could be accomplished by placing all possible values on a scale indicating how close to the desired value they are. Finding a way to assign a universally approved score to each criterion is outside the scope of my thesis so below we just present our design choice. In order to find a universally approved way of assigning scores to the criteria we suggest a case study to identify different groups of users and their opinion on the matter (like we did with the user group analysis).
b. At which point should quantification of each criterion occur:
To determine when the quantification of each criterion should occur the list of criteria has to be separated in two categories. One category contains the criteria that are purpose or query dependent which means that the desired value for them is determined only after the user has inserted the purpose of search or query. The

other category contains non-purpose, non-query dependent criteria, therefore criteria for which the desired value is constant no matter the purpose or query of search. Below the two categories are shown:

1. Non-purpose/Non-query dependent criteria:
   - Technical Characteristics:
     No matter what the query or purpose of search is, educational material with higher resolution are always better.
   - Copyright License:
     Irrespectively of the query or purpose of search the more permissions granted by a copyright license the better. The desired value for this criterion is the most open license according to Creative Common's licensing spectrum (see appendix A).
   - Editable:
     Regardless of the query or purpose of search, material that have no logos or watermarks are always preferable to ones filled with them.
   - Well-Organized, Comprehensible and Well-presented (proxied with rating scores):
     We want to get the highest possible rating for an educational material no matter what our purpose or query is.
   - Number of topics:
     A resource that focuses exclusively on the query keyword as a topic is always preferable to resources that concern a multitude of topics.
2. Query/Purpose dependent criteria:
   - Relevancy:
     This criterion exists solely in relation to the keyword used for a search therefore it is query dependent. However regardless of the purpose or query we want to get educational material with the highest relevancy score available.
   - Up-to date:
     Irrespectively of the purpose of search we want to get material that were created recently, instead of getting outdated ones. However the score we assign is dependent on the topic of query. If a topic is well established and has been fully developed for years perhaps all relevant material will be out-dated (this decision was made based on insights we got from the user-group analysis). Therefore assigning the score should be done according to the query.
   - Length:
     The desired value for this criterion is influenced by the purpose of search. More specifically, if the users are looking for material to prepare a lecture presentation then they want short length concise content. However if they are trying to expand their knowledge a lengthy detailed description is preferable in order to understand the content.
   - Educational level:
     Considering all possible purposes for searching material it is obvious that the use of this criterion is purpose dependent. For example if the purpose of search was **Researching in order to remember prior knowledge** making use of the educational level criterion does not make sense because all educational level materials could be useful for the user.
   - Reliable (proxied with institution's/platform's/author's reputation):
     This criterion is also query dependent since the credibility of a source is always in relation to the topic of discussion. Whereas one organization might be considered an authority for its department in physics that is not necessarily true for its other departments. So the desired value for this criterion is decided based on the query.

From the list above it is obvious that for some criteria their score can be decided prior to any queries, however there are some that depend on the users' input and thus can only be decided on the spot. So we have four alternatives for when and where to assign the individual scores:

1. **First option:** All scores can be assigned after the retrieval of all the relevant resources of a query. This choice however could have a negative impact on the efficiency of the model due to delays for the calculation of these scores.

2. **Second option:** All scores can be assigned to all criteria before any queries are made. This means sweeping assumptions would be made regarding the desired value for the purpose/query dependent criteria. This could result in the assignment of the wrong score to some criteria and therefore mistakes in the calculation of the quality score of some resources.

3. **Third option:** All the non-purpose/non-query dependent criteria can be assigned scores prior to any queries. Whereas all the query or purpose dependent ones can be assigned scores after a query is made. This choice will most likely give the most accurate quality score but will influence negatively the efficiency of the model.

4. **Fourth option:** Assign scores only to the non-purpose/non-query dependent criteria and ignore the query or purpose dependent ones. This choice disregards the findings of the literature survey regarding the criteria to quantify quality and provides a quality score without making full use of all the available information of a resource.

Design Choices:
a. How to assign an individual quality score to each criterion:
For my implementation the score assignment for each criterion will be done on a scale from 0 to 1. Where 1 is the highest value that could be assigned to a criterion and 0 represents the absence of information regarding a criterion. This decision was made based on literature findings which say the absence or bad quality of the metadata of a resource is an indication of low quality OERs.

Unfortunately, the metadata offered from the repositories used in our implementation do not contain information about all the criteria mentioned in the list above. So for my implementation a subset of these criteria is used. Below I present how the individual score assignment of each criterion is accomplished in our implementation of this quality model.

1. Non-purpose/non-query dependent criteria:

    - Technical Characteristics:
      This criterion is only applicable for video and image resources. The individual quality scores are assigned as follows: 3840x2160 or 2160p =1, 2560x1440 or 1440p=0.9, 1920x1080 or 1080p =0.8,1280x720 or 720p=0.7, 854x480 or 480p=0.5, 640x360 or 360p=0.3 and 426x240or 240p=0.1.

    - Copyright License:
      Using the categorization of creative commons copyright licenses the scores are assigned like this: CC0=1, CC-BY=0.9, CC-BY-SA=0.7, CC-BY-NC=0.6, CC-BY-NC-SA=0.5, CC-BY-ND=0.3,CC-BY-NC-ND=0.2, C=0.1.

    - Rating (Well-organized, Comprehensible, Well-presented):
      5 stars=1, 4 stars=0.8, 3 stars=0.6,2 stars=0.4, 1 star =0.2. Any rating value that includes decimals is divided by 5.

    - Number of topics:
      For this criterion the individual quality score is assigned as follows: 1 topic = 1, 2 topics = 0.95, 3 topics = 0.85, 4 topics=0.75, 5 topics = 0.6, 6 topics = 0.5, 7 topics = 0.4, 8 topics = 0.3, 9 topics = 0.2 and anything with 10 or more topics gets a score of 0.1.

2. Query/Purpose dependent criteria:

    - Up-to date:
      For this criterion after retrieving all relevant results they are sorted based on their creation date starting with the most recent. Then the scores are assigned accordingly. Specifically we divide the sorted list in 20 groups and each group is assigned a value from 1 to 0.05. The score assigned to each successive group is assigned a value decreased by 0.05.

      Alternative solution: The scores regarding this criterion could also be assigned based on a threshold, meaning find the current date and anything that was created in the past year gets a score of 1 anything between two years and one year prior gets a score of 0.9 etc. We chose not to follow this implementation due to the fact there are topics for which no recent resources exist which means all resources could be assigned a low score.

- Number of pages (Length):
  The desired value for this criterion changes according to the purpose of search. The score for this criterion is assigned in the same way as the score for the criterion up-to-date is assigned. However the difference is that the sorting changes based on the purpose of search. For the purpose **Preparing a presentation for a lecture** we sort in **ascending order** from smallest to largest. For the purpose **Expanding knowledge** we sort the resources in **descending order** from largest to smallest.

  Alternative: Use threshold length to assign scores. The drawbacks of this approach are the same as with the up-to-date criterion since all resources may have a small or large number of pages which makes finding a threshold difficult.
- Source (Reliable):
  In order to assign a score to this criterion ranked lists of all available organizations on a variety of topics were needed. In our implementation all resources originate from two Universities MIT and TU Delft. So for our implementation we decided to use the Shanghai Academic Ranking of World Universities to find in which place these two Universities are ranked for different fields of study and assign scores to them accordingly. So if for example a resource was created by someone affiliated with MIT and MIT was placed in the first 50 places on the ShanghaiRanking for that topic then 1 is assigned to the source of that resource.

b. When and where should the scores be assigned:
Regarding the matter of when and where the individual criterion scores should be assigned we implemented the first option because we did not have access to the database of resources in order to implement the third option. We think that the third option would yield the most accurate quality scores for the resources and will not affect the efficiency of the model. In our implementation we managed to assign the non-purpose/non-query dependent criteria with computational complexity O(1). So we managed to approximate the third option we presented.

5.3.4. STEP 4

The fourth step consists of assigning weights to each criterion to indicate their importance in calculating the quality of educational resources. At this point we need to define a preference function for the criteria, meaning we need to decide a) how and b) when to distribute the weights among the criteria.

This step caters to the relative facet of quality. However it is impossible to find a preference function that matches everybody's perception of what constitutes high quality educational material. That is why the users are asked to insert their preferences. From the user group analysis we were also able to identify two sets of weights that show our target audience's perception of quality. The user group analysis also showed that this preference function depends on the purpose of the search. Below we present the design space, my design choice and some alternatives.

Design Space:

a. How to distribute the weights amongst the quality criteria:
As with the individual scores, there is an infinite number of ranges of values and combinations for the assignment of weights to the quality criteria.

b. Ways of assigning weights:
Similarly with the assignment of individual scores to the criteria there is a challenge regarding when these weights should be assigned to each criterion. Below we describe the five possible ways of achieving this:

1. **First option:** The first option consists in the assignment of fixed weights to the criteria irrespectively of the purpose of search and user. This option is not ideal because it is based on assumptions regarding the preferences of users and disregards the differing perceptions regarding quality. In addition, it contradicts the findings of the user group analysis that showed difference in perception based on the purpose of search.

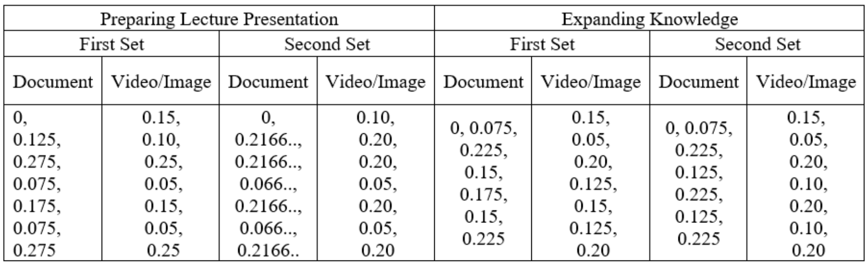
2. **Second option:** This option suggests that all weights be assigned by the user before making the query. While this way caters to the subjective, ever-changing nature of quality it is not practical in application.

3. **Third option:** Some of the weights are assigned beforehand, more specifically the criteria identified as less important from the user group analysis. As with the first option, assumptions are made about how people perceive quality and this might negatively impact the quality quantification.

4. **Fourth option:** Provide multiple sets of weights for each purpose and allow the user to select one of them.

5. **Fifth option:** Provide the users with one default set of weights for each purpose but also allow them to change all attributes at will.

Design Choices:
a. How to distribute the weights among the quality criteria:
In my implementation the weights per purpose are assigned based on a qualitative assessment of the ordering of the criteria by the participants of the user group analysis and the results of a quality panel organized by TU Delft's library (see section 4). From the user group analysis and the quality panel we were able to identify two sets of weights. The analysis also indicated this should change according to the purpose of search. In addition, the list of criteria we use contains the technical characteristics criterion which can only be applied to video/image resources so we need to adjust the weights according to the type of resource as well. In figure 5.2 we present the distribution of the quality weights we used in our implementation. Unfortunately, we were not given access to video resources so those sets of weights were never used.



Preparing Lecture Presentation				Expanding Knowledge			
First Set		Second Set		First Set		Second Set	
Document	Video/Image	Document	Video/Image	Document	Video/Image	Document	Video/Image
0, 0.125, 0.275, 0.075, 0.175, 0.075, 0.275	0.15, 0.10, 0.25, 0.05, 0.15, 0.05, 0.25	0, 0.2166.., 0.2166.., 0.066.., 0.2166.., 0.066.., 0.2166..	0.10, 0.20, 0.20, 0.05, 0.20, 0.05, 0.20	0, 0.075, 0.225, 0.15, 0.175, 0.15, 0.225	0.15, 0.05, 0.20, 0.125, 0.15, 0.125, 0.20	0, 0.075, 0.225, 0.125, 0.225, 0.125, 0.225	0.15, 0.05, 0.20, 0.10, 0.20, 0.10, 0.20

Figure 5.2: Figure that shows the distribution of weights for all possible scenarios, the weight are presented in the following order **Resolution, Copyright License, Rating, Number of topics, Up-to-date, Length, Source**

b. Ways of assigning weights:
When implementing this quality model we chose to offer the users two sets of weights as well as the option to select their own weights of importance. For our implementation we combined the fourth and fifth options because those are the two options that better cater to the subjective nature of the concept of quality. The reason we offer two possible sets of weights for each purpose is because we were able to identify that many differing views from the qualitative assessment of the user group analysis and the panel on quality organized by the library.

5.3.5. STEP 5

At this point each quality criterion has a score and except for relevancy a weight of importance. Disregarding relevancy for the moment a way to combine all these scores and weights is needed in order to calculate an intermediate quality score. Relevancy should not be included in this calculation because it is the most important criterion for quality (this is evidenced by the fact it is the only criterion used by current search engines). Therefore this intermediate score represents all the other criteria that are not being used by current search engines and which should be treated separately. In order to combine these criteria and quantify quality some kind of mathematical formula is needed.

Design Space:
The design space for the calculation of this intermediate quality score consists of all conceivable formulas that help combine a number of values with weights.

Design Choice:
My design choice for quantifying the quality of an educational resource based on this list of criteria relies on the use of the following formula:

$$Q = \sum_{i=1}^{n} w_i \times crit_i$$

where
Q: is the intermediate quality score assigned to a resource, this score takes into account all the quality criteria except relevancy, $0 \leq Q \leq 1$
n: is the number of criteria in the set of criteria determined by the purpose of search, type of resource and

used to quantify quality
w_i: is the weight assigned to criterion i, $0 \leq w_i \leq 1$ and the sum of all the weights must be equal to 1

$$\sum_{i=1}^{n} w_i = 1$$

$crit_i$: is the quality value assigned to criterion i, $0 \leq crit_i \leq 1$

My decision to use this formula for the calculation of quality was influenced by my research on quality models where I encountered similar formulas [17] and because it is the natural choice when trying to combine scores with weights. The use of weights reflects the subjectivity of quality and the change in importance given to each criterion depending on the purpose of use and the users' individual preferences. In addition, the criteria used to measure quality are ever-changing so there is a need for a formula that is flexible and can be easily adjusted. The proposed formula provides all these since criteria can easily be removed or added by simply changing the value of n and adjusting the weights accordingly.

5.3.6. STEP 6

The final step of the quality model consists of the combination of the relevancy, the intermediate quality score and the fitness to purpose parameter in order to calculate a final quality score (Final_Q). Below I explain in detail why these three values are needed to achieve quantification of the quality of educational resources.

Explanation of the values that need to be combined for quality quantification:

Relevancy: The relevancy score of a resource shows how relevant a resource is to the query made by the user. Therefore an irrelevant resource is useless to the user and thus will be perceived as low quality.

Intermediate quality score: This score combines all the criteria that color the users' perception regarding quality except for relevancy. These are additional indicators of the usefulness of a resource.

Fitness to purpose: From the user group analysis it is apparent that based on the purpose of search some materials are considered of high usefulness while others less so. The use of this parameter caters to the purpose-dependency facet of quality and is another consideration that needs to be taken into account when calculating the final quality score.

We hypothesize that when these three values are combined the results of a resource can be re-ranked in a more efficient way.

Design Space:
The design space of the final step of the quality model consists of the calculation of a) the fitness to purpose parameter and b) the final quality score.

a. Calculation of Fitness to Purpose parameter:
This parameter indicates whether the type of a resource is useful based on the purpose of search. For example the user group analysis (see chapter 4) showed that for the purpose of **Preparing a lecture presentation** lecture slides, videos and images are the most useful type of resources. The assignment of a score to this parameter requires a knowledge of all available type of materials and a categorization of these regarding their usefulness.

b. Calculation of final quality score:
The design space for the calculation of the final quality score consists of all possible formulas that can combine the three components: relevancy, intermediate quality score and fitness to purpose parameter. Since the final goal is to re-rank the results based on this final quality score the first thing to consider is how the re-ranking should look in the end. Below we present how this formula should function for all the possible scenarios:

Re-ranking based on relevancy and intermediate quality score:

1. Low relevancy score + Low quality score (Q) = Low final quality score (Final_Q): Educational resources with such a result should be at the bottom of the list after re-ranking.

2. Low relevancy score + Medium quality score (Q) = Low final quality score (Final_Q): Educational resources with such a result should be at the bottom of the list after re-ranking.

3. Low relevancy score + High quality score (Q) = Low final quality score (Final_Q): Educational resources with such a result should be at the bottom of the list after re-ranking.

4. Medium relevancy score + Low quality score (Q) = Low final quality score (Final_Q): Educational resources with such a result should be at the bottom of the list after re-ranking.

5. Medium relevancy score + Medium quality score(Q) = Medium final quality score (Final_Q): Material with such a result should be at the middle of the list after re-ranking.

6. Medium relevancy score + High quality score (Q) = Medium final quality score (Final _Q): Resources with such a result should be at the middle of the list after re-ranking.

7. High relevancy score + Low quality score (Q) = Medium final quality score (Final_Q): Material with such a result should be at the middle of the list after re-ranking.

8. High relevancy score + Medium quality score (Q) = High final quality score (Final_Q): Educational resources with such a result should be at the top of the list after re-ranking.

9. High relevancy score + High quality score (Q) = High final quality score (Final_Q): Material with such a result should be at the top of the list

Re-ranking based on fitness to purpose parameter:

1. If the fitness to purpose parameter of a resource is equal to 1, irrespective of the combination of relevancy and intermediate quality it should be placed at the top of the list.

2. If the fitness to purpose parameter of a resource is equal to 0.5, no matter the combination of relevancy and intermediate quality it should be placed in the middle of the list of available resources.

3. If the fitness to purpose parameter is equal to 0 then this resource should be placed at the bottom of the list.

Figure 5.3 shows the order in which these results should be placed in the final re-ranked list presented to the user.

Some overlap between neighbouring scores is permitted e.g resources with medium final score could potentially be placed among high final scores. However resources with low relevancy scores should never be placed among resources with medium or high final quality scores. This design decision is based on the fact that users sometimes have to compromise if they do not find exactly what they are looking for (meaning in the absence of resources with high relevancy and high quality).

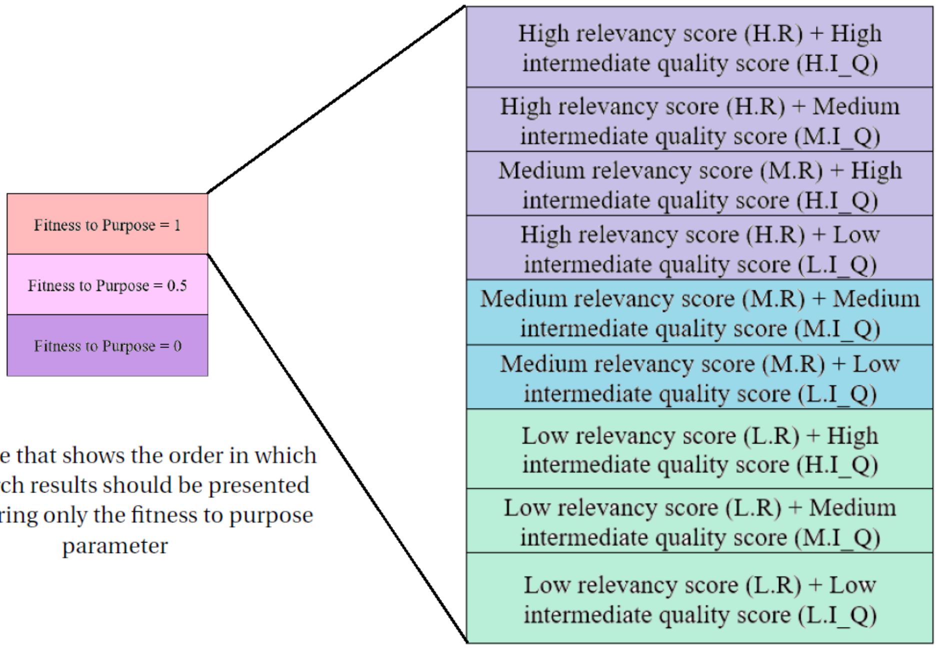
Design Choices:
a. Calculation of Fitness to Purpose parameter:
In our implementation the resources come from two repositories MIT open courseware and TU Delft's educational and research repositories. The MIT repository is filled with course materials used during lectures. The educational repository of TU Delft is filled with students' theses, while the educational repository of TU Delft is filled with journal articles, conference papers, books, book chapters etc. So based on the purpose of search and the repository from which a resource originated the fitness to purpose parameter is calculated.

Preparing a lecture presentation:

- MIT repository: Fitness to purpose = 1
- Research repository of TU Delft: Fitness to purpose = 0.5
- Educational repository of TU Delft: Fitness to purpose = 0

Expanding knowledge:

- MIT repository: Fitness to purpose = 0.5


Fitness to Purpose = 1
Fitness to Purpose = 0.5
Fitness to Purpose = 0
High relevancy score (H.R) + High intermediate quality score (H.I_Q)
High relevancy score (H.R) + Medium intermediate quality score (M.I_Q)
Medium relevancy score (M.R) + High intermediate quality score (H.I_Q)
High relevancy score (H.R) + Low intermediate quality score (L.I_Q)
Medium relevancy score (M.R) + Medium intermediate quality score (M.I_Q)
Medium relevancy score (M.R) + Low intermediate quality score (L.I_Q)
Low relevancy score (L.R) + High intermediate quality score (H.I_Q)
Low relevancy score (L.R) + Medium intermediate quality score (M.I_Q)
Low relevancy score (L.R) + Low intermediate quality score (L.I_Q)

(a) Figure that shows the order in which the search results should be presented considering only the fitness to purpose parameter

(b) Figure that shows the order in which the search results should be presented considering only relevancy and the intermediate quality score

Figure 5.3: Re-rankings of search results based on the three components: Relevancy, Intermediate Quality score and Fitness to purpose parameter

- Research repository of TU Delft: Fitness to purpose = 1
- Educational repository of TU Delft: Fitness to purpose = 0

b. Calculation of final quality score: For the calculation of the final quality score we came up with four different mathematical formulas. Below we present each formula:

1. **Option 1 (Weighted_Sum):**

$$Final\_Q = fitness \times (0.7 \times Relevancy\_Score + 0.3 \times Q)$$

   where
   Final_Q: final quality score that includes relevancy, $0 \leq Final\_Q \leq 1$
   Relevancy_Score: relevancy of the resource to the query generated by the search algorithm and normalized so that its range is $0 \leq Relevancy\_Score \leq 1$
   Q: quality score of all quality attributes except relevancy calculated by the formula presented in section *Point 5*, $0 \leq Q \leq 1$
   fitness: the fitness to purpose parameter, $0 \leq fitness \leq 1$

   The first option is a weighted sum of the relevancy and intermediate quality score of each resource where a greater weight is assigned to relevancy. The combination of these two components is then multiplied by the fitness to purpose parameter so that resources are ordered based on how well they serve the purpose of search (see figure 5.3).

2. **Option 2 (Mult):**

$$Final\_Q = fitness \times (Relevancy\_Score \times Q)$$

   In this formula instead of assigning a greater weight of importance to one of the two components relevancy and intermediate quality score, they are treated equally. Then the fitness to purpose parameter is used to sort the resources based on their usefulness to the purpose.

3. **Option 3: (Mult_Sqr)**

$$Final\_Q = fitness \times (Relevancy\_Score^2 \times Q)$$

Similarly to the previous formula, in this one the relevancy and intermediate quality score are also multiplied. However the relevancy score is given a greater importance to make sure that resources with the lower relevancy scores will be placed at the bottom of the list.

4. **Option 4 (Branched):**

$$Final\_Q = \begin{cases} fitness \times Relevancy\_Score^2 \times Q, & \text{if L.R} \\ fitness \times Relevancy\_Score \times Q, & \text{if M.R \& L.I\_Q} \\ fitness \times (0.7 \times Relevancy\_Score + 0.3 \times Q), & \text{if M.R \& H.I\_Q, H.R \& L.I\_Q} \\ fitness \times (0.8 \times Relevancy\_Score + 0.2 \times Q), & \text{otherwise} \end{cases}$$

where
Low Relevancy (L.R): $L.R < 0.75$,
Medium Relevancy (M.R): $0.75 \leq M.R < 0.9$,
High Relevancy (H.R): $H.R \geq 0.9$,
Low Intermediate Quality (L.I_Q): $\text{L.I\_Q} < 0.3$,
Medium Intermediate Quality (M.I_Q): $0.3 \leq \text{M.I\_Q} < 0.5$,
High Intermediate Quality (H.I_Q): $\text{H.I\_Q} \geq 0.5$

This mathematical formula is a combination of the previously presented options. Here a different equation is used depending on the combination of relevancy and intermediate quality scores. When a resource has high relevancy a weighted sum is calculated with greater weight to relevancy in order to show its importance in determining quality. When the relevancy is medium and the quality high or we have high relevancy and low quality a different weighted sum is calculated with a slightly smaller weight to relevancy and higher weight to quality. When resource has medium relevancy and low quality we multiply the two scores (relevancy and quality) in order to get a low final score. Such resources should be placed near the bottom of the list. For the same reason when a resource has a low relevancy we square it and multiply with the intermediate quality score. In all the cases the combination of relevancy and intermediate quality are multiplied by the fitness to purpose parameter to achieve the ordering presented in figure 5.3.a.

When implementing the quality model all four mathematical formulas were used. To select the formula that best re-ranks the resources to the satisfaction of the users, experiments were conducted (see chapter 7).

6

Implementation

In this chapter we present the implementation of a prototype search tool that integrates the quality model presented in the diagram of section 5.2. Using the diagram in figure 5.1 as a guide and implementing the design choices presented in the previous chapter we were able to build a prototype search tool that re-ranks educational resources based on their quality. Below we give more details on how this tool was developed, we show an empathy map that was used to ascertain all the functionalities needed and some screen shots of this tool.

6.1. Technical Details

In this section we present the various components used to develop our prototype search tool which is capable of calculating and re-ranking the results of a search query.

6.1.1. Search Engine

In order to implement the quality model presented in figure 5.1 a search engine was required. For this implementation, the company Feedback Fruits lent us their prototype search engine. This search engine works with a customized version of the Okapi BM25 algorithm and the relevancy of each resource is determined based on a topic extraction from its title, keywords and abstract, meaning only topical relevancy is calculated. Furthermore, this prototype search engine was developed in *TypeScript* with the help of the following components:

- Elastic Search (an open source search engine built on top of Apache Lucene).
- Apache Kafka (a distributed streaming platform)
- GraphQL (query language for executing database queries)

6.1.2. Resources

The educational material used in our implementation originate from the MIT open courseware and the TU Delft repositories. The MIT open courseware repository offers course materials in the form of document and video resources. Whereas TU Delft has two separate repositories: a) the Educational repository which offers bachelor and master theses of its students and b) the Research repository which offers journal articles, conference papers, reviews, reports, books, book chapters and public lectures.

To quantify quality we used all the available metadata tags of these resources and if some needed information was missing we extracted it (see section 3.4). For example all resources from both repositories did not include information regarding the number of pages of a resource, so we processed them to get that information. The MIT resources did not include information regarding the creation date so we processed them as well. Additionally, resources from the TU Delft repository required extra processing so that they could be integrated in the search engine of Feedback Fruits. Specifically, their format had to be changed so they could be processed through Feedback Fruit's system.

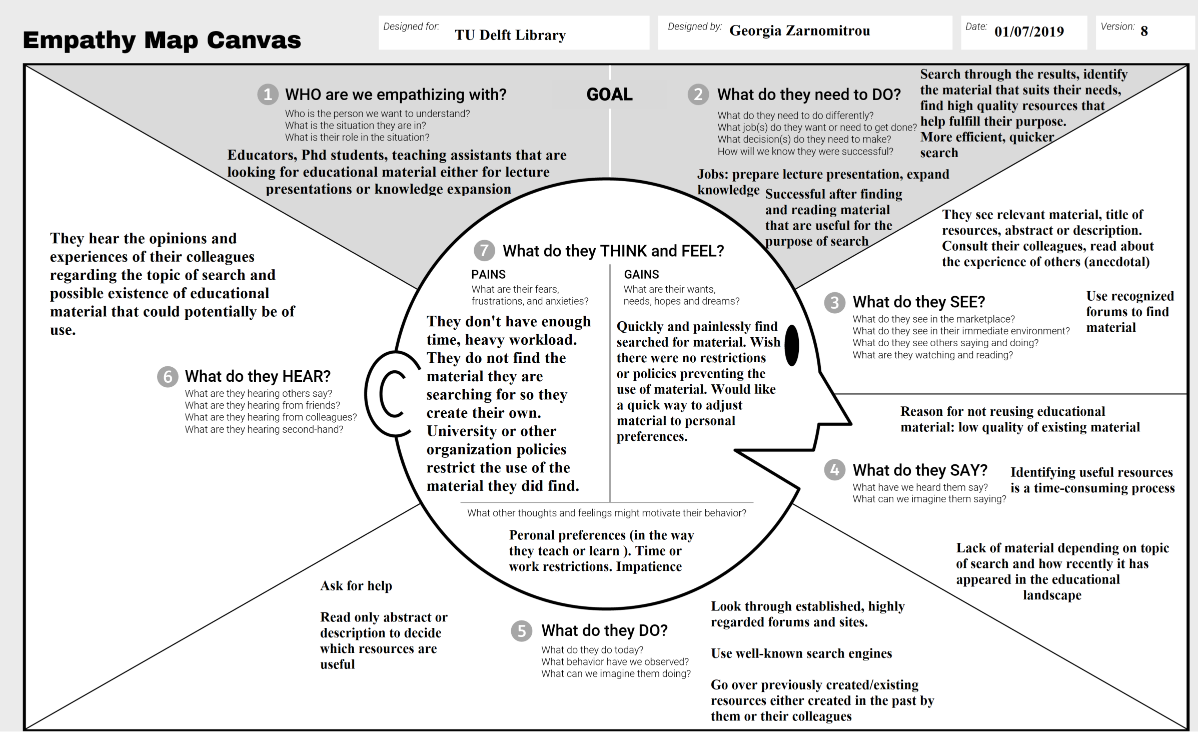
6.1.3. TOOLS USED FOR IMPLEMENTATION

This prototype search tool that automatically calculates the quality of resources and re-ranks them accordingly was implemented with *Javascript*. For the implementation of our quality model and the interface we used the following components:

- GraphQL: The use of this query language is necessary because Feedback Fruits provided us a GraphQL access point in order to communicate with their database. So in this tool GraphQL is used to retrieve any resources that are relevant to the query of a user. Along with the resources, the GraphQL access point provides all the available metadata and the relevancy scores of these resources.
- Node.js: This open source platform is used to run the program
- React: This javascript library is used to build the interface of the tool.

6.2. FUNCTIONALITIES OF THE PROTOTYPE TOOL

During the development of this prototype tool we used the information gathered from the user group analysis to create an empathy map. This map allowed us to gain a deeper insight in the thoughts and feelings of the users and corroborate that the functionalities offered in our implementation serve their purpose. The empathy map created for the implementation of this prototype is shown in figure 6.1.


Empathy Map Canvas
Designed for:
Designed by:
Date:
Version:
GOAL
1 WHO are we empathizing with?
Who is the person we want to understand?
What is the situation they are in?
What is their role in the situation?
2 What do they need to DO?
What do they need to do differently?
What job(s) do they want or need to get done?
What decision(s) do they need to make?
How will we know they were successful?
3 What do they SEE?
What do they see in the marketplace?
What do they see in their immediate environment?
What do they see others saying and doing?
What are they watching and reading?
4 What do they SAY?
What have we heard them say?
What can we imagine them saying?
5 What do they DO?
What do they do today?
What behavior have we observed?
What can we imagine them doing?
6 What do they HEAR?
What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second-hand?
7 What do they THINK and FEEL?
PAINS
What are their fears, frustrations, and anxieties?
GAINS
What are their wants, needs, hopes and dreams?
What other thoughts and feelings might motivate their behavior?

Designed for: TU Delft Library

Designed by: Georgia Zarnomitrou

Date: 01/07/2019

Version: 8

Educators, Phd students, teaching assistants that are looking for educational material either for lecture presentations or knowledge expansion

Search through the results, identify the material that suits their needs, find high quality resources that help fulfill their purpose. More efficient, quicker search

Jobs: prepare lecture presentation, expand knowledge

Successful after finding and reading material that are useful for the purpose of search

They see relevant material, title of resources, abstract or description. Consult their colleagues, read about the experience of others (anecdotal)

Use recognized forums to find material

Reason for not reusing educational material: low quality of existing material

Identifying useful resources is a time-consuming process

Lack of material depending on topic of search and how recently it has appeared in the educational landscape

Look through established, highly regarded forums and sites.

Use well-known search engines

Go over previously created/existing resources either created in the past by them or their colleagues

Ask for help

Read only abstract or description to decide which resources are useful

They hear the opinions and experiences of their colleagues regarding the topic of search and possible existence of educational material that could potentially be of use.

They don't have enough time, heavy workload. They do not find the material they are searching for so they create their own. University or other organization policies restrict the use of the material they did find.

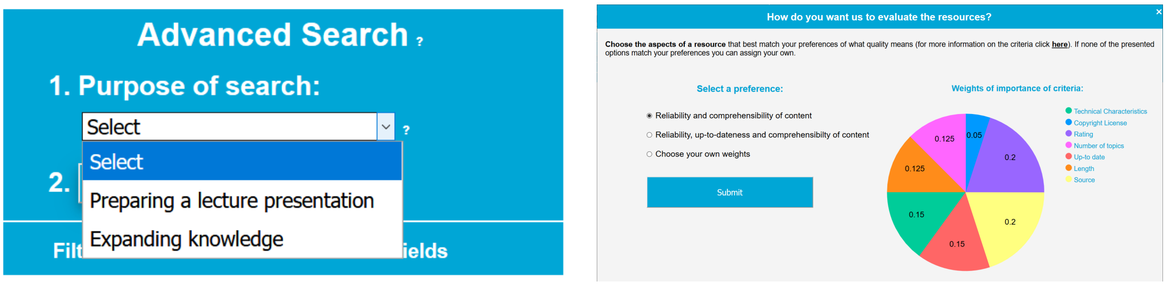
Quickly and painlessly find searched for material. Wish there were no restrictions or policies preventing the use of material. Would like a quick way to adjust material to personal preferences.

Peronal preferences (in the way they teach or learn). Time or work restrictions. Impatience

Figure 6.1: Empathy map that depicts the feelings, thoughts and actions of users who interact with a search engine

Below we present several screen shots of the various functionalities offered by this prototype tool.
Input of purpose of search and personal preferences: Since quality is purpose dependent the tool offers the user a list of educational goals so the quality quantification can be adjusted based on it. Besides the purpose of search the user is asked to insert his perception of quality by selecting one of the different sets of weights presented or by creating his own set of weights. These functionalities are presented in figure 6.2.

Add Resource Functionality: The empathy map (see figure 6.1) shows the users' frustration with current University and organization policies that impose restrictions on the access and use of educational resources. In addition, from our literature survey and user group analysis it is apparent that some educators perceive Open Educational Resources as low quality material. Therefore we propose the **Add Resource** functionality that allows users to freely share any resources they create. This way the quality of the available resources will improve and there will be no restrictions regarding reuse. Furthermore this feature ensures all needed metadata



How do you want us to evaluate the resources?

Choose the aspects of a resource that best match your preferences of what quality means (for more information on the criteria click **here**). If none of the presented options match your preferences you can assign your own.

Select a preference:

● Reliability and comprehensibility of content

○ Reliability, up-to-dateness and comprehensibilty of content

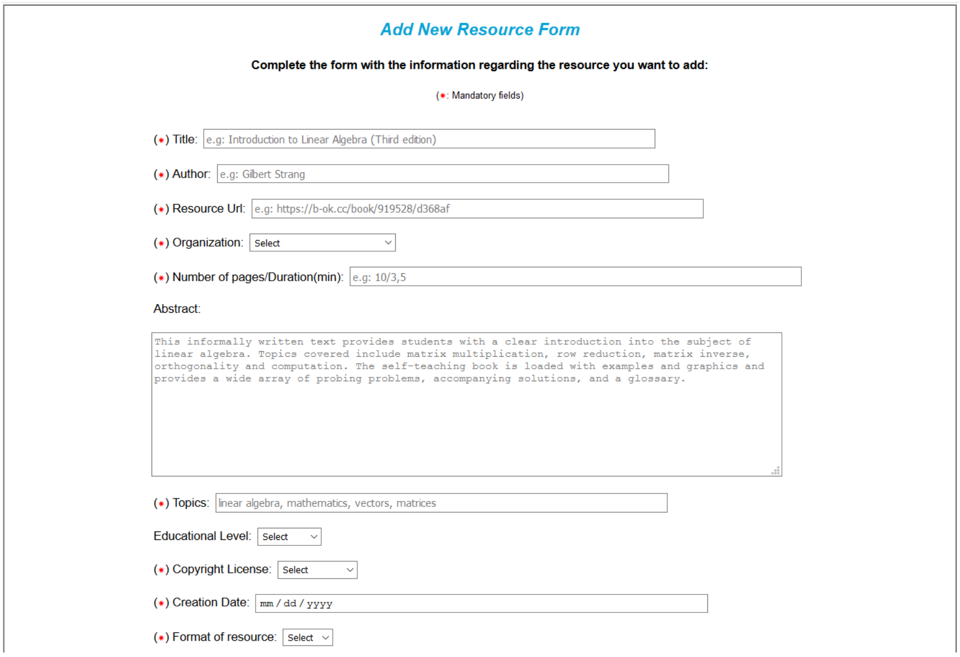
○ Choose your own weights

(a) Functionality to insert purpose of search

(b) Functionality to insert user preferences

Figure 6.2: Functionalities for user input

to quantify quality are inserted by the user. This functionality is shown in figure 6.3.



Add New Resource Form

Complete the form with the information regarding the resource you want to add:

(*: Mandatory fields)

(*) Title: e.g: Introduction to Linear Algebra (Third edition)

(*) Author: e.g: Gilbert Strang

(*) Resource Url: e.g: https://b-ok.cc/book/919528/d368af

(*) Organization: Select

(*) Number of pages/Duration(min): e.g: 10/3,5

Abstract:

```
This informally written text provides students with a clear introduction into the subject of
linear algebra. Topics covered include matrix multiplication, row reduction, matrix inverse,
orthogonality and computation. The self-teaching book is loaded with examples and graphics and
provides a wide array of probing problems, accompanying solutions, and a glossary.
```

(*) Topics: linear algebra, mathematics, vectors, matrices

Educational Level: Select

(*) Copyright License: Select

(*) Creation Date: mm / dd / yyyy

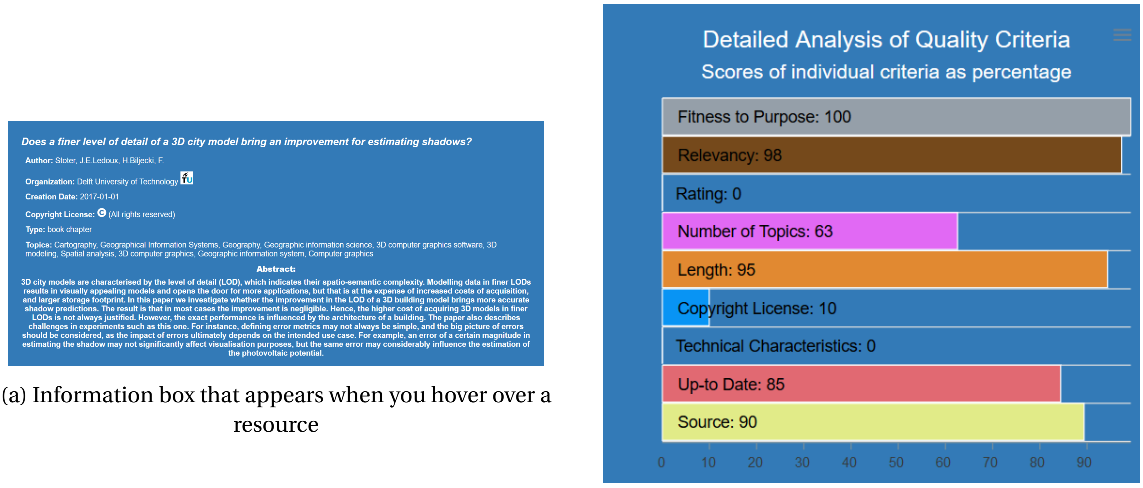
(*) Format of resource: Select

Figure 6.3: Functionality to add new resource

Functionalities to provide more information about each resource: From the empathy map it is also clear that searching for appropriate high quality educational material is a time-consuming process. During the user group analysis several interviewees mentioned that they sometimes identify high quality material from the description or abstract. Therefore in our implementation we offer two features that present pertinent information about each resource, an information box when you hover over the title of a resource and a detailed analysis of the various quality criteria scores. These features can be seen in figure 6.4.

Visualization: In this tool we also offer a visualization feature. Specifically, we visualize the up-to-dateness of the search results we get from a query through a bar chart. This bar chart shows the years in which the resulting resources were created and how many resources were created each year. We also offer a word cloud that shows all related keywords and topics to a search query. This visualization is useful for recognizing topics that are related to the query keyword that was used. In the future if these features were made interactive (meaning users could access resources through it), they would prove very useful in further assisting educators in narrowing down their search and browsing for needed material. In figure 6.5 we can see a screen shot of these features.

This implementation is only a prototype, however in the future there are many additional functionalities and features that could be added to cater to the needs of the users. For example the addition of a comment



Does a finer level of detail of a 3D city model bring an improvement for estimating shadows?

Author: Stoter, J.E.Ledoux, H.Biljecki, F.

Organization: Delft University of Technology

Creation Date: 2017-01-01

Copyright License: (All rights reserved)

Type: book chapter

Topics: Cartography, Geographical Information Systems, Geography, Geographic information science, 3D computer graphics software, 3D modeling, Spatial analysis, 3D computer graphics, Geographic information system, Computer graphics

Abstract:

3D city models are characterised by the level of detail (LOD), which indicates their spatio-semantic complexity. Modelling data in finer LODs results in visually appealing models and opens the door for more applications, but that is at the expense of increased costs of acquisition, and larger storage footprint. In this paper we investigate whether the improvement in the LOD of a 3D building model brings more accurate shadow predictions. The result is that in most cases the improvement is negligible. Hence, the higher cost of acquiring 3D models in finer LODs is not always justified. However, the exact performance is influenced by the architecture of a building. The paper also describes challenges in experiments such as this one. For instance, defining error metrics may not always be simple, and the big picture of errors should be considered, as the impact of errors ultimately depends on the intended use case. For example, an error of a certain magnitude in estimating the shadow may not significantly affect visualisation purposes, but the same error may considerably influence the estimation of the photovoltaic potential.

(a) Information box that appears when you hover over a resource

(b) Score break down analysis which appears when you hover over the quality score of a resource

Figure 6.4: Functionalities for user input

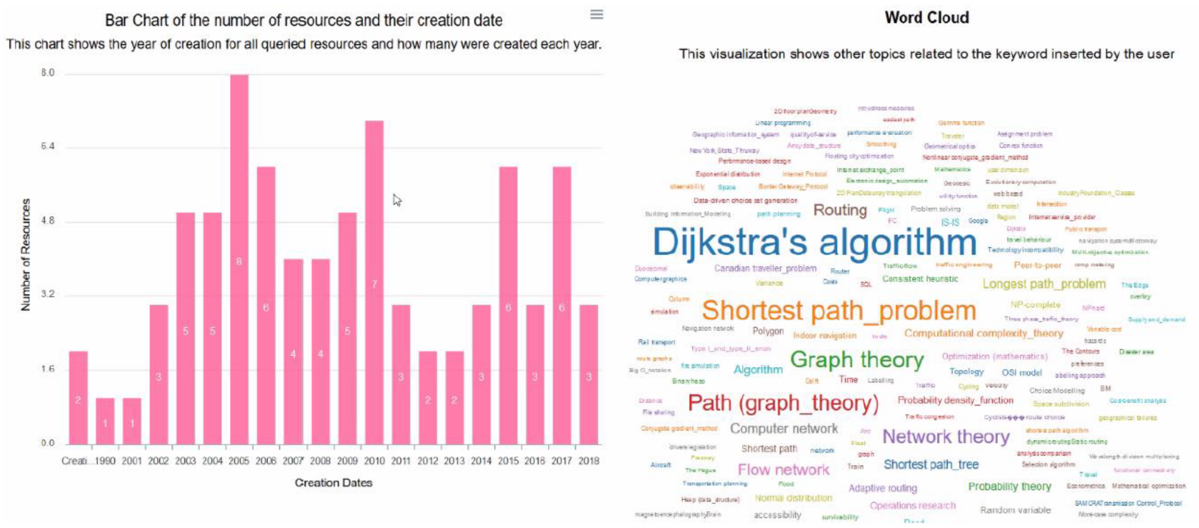


Figure 6.5: Visualization feature

section or a forum where educators can exchange ideas and experiences would be a good idea. Another useful addition could be a recommendation system that provides the user with helpful suggestions of resources they might be interested in. These kind of functionalities could also contribute to the calculation of the quality of resources e.g sentiment analysis of comment section, personalization through recommendations (users' view history) etc.

6.3. RESULT

From this implementation the main take away is that educational repositories have really bad metadata tagging practices. This was evident by the fact the majority of time it took to implement this model was spent in trying to extract needed information from the resources or re-process them. In addition, the metadata information that is provided from educational repositories is not rich enough to allow us to use all the criteria we identified from our literature survey. Furthermore, the metadata available from the repository of MIT was really sparse. From this implementation we also noticed that due to the way we assign scores to the individual quality criteria the automatic assessment of quality causes delays. Specifically, the sorting operations done based on the creation date and length of the resources might cause some delay depending on the number of results we get.

With the completion of this implementation we achieved a number of tasks. We implemented the quality model we designed (see chapter 5) and achieved to automatically assess the quality of resources that result from a search query. This responds to our second research question on whether automatic quality assessment is possible. This implementation proves that indeed we can automatically calculate a quality score for each resource. However, to see if the score we calculate really represents a significant quality quantification we need to test if its integration in the ranking of the results makes a difference. So, this quality quantification is used to offer different re-rankings of the search results. Specifically, this tool ranks search results in five different ways: based on the topical relevancy of the resources (search algorithm used by Feedback Fruits) and based on the four re-rankings we propose in section 5.3.6. Therefore, we have a tool that allows us to view and compare these five different rankings of search results. In the following chapter we present the results of an experiment we conducted to test whether our quality quantification is significant and thus improves the re-ranking of the results. With this evaluation we answer the remaining research question of this thesis (if the automatic quality assessment improves the ordering of current search algorithms).

7 Evaluation

In this chapter, we present the evaluation of the quality model we designed and implemented. Specifically, using the prototype tool presented in the previous chapter (see chapter 6) we conducted an experiment to test whether the automated assessment of the quality of educational material can be used to enhance current search processes. This will help us in answering our final research question.

7.1. Purpose of Analysis

After implementing the quality model presented in chapter 5, we wanted to test whether the automated quality quantification we propose can help improve the ranking offered by current search engines. With the results of this test we will ascertain whether an improvement of the ranking of search results is possible. With this experiment we will also have proof on whether we are successful in automatically assessing the quality of educational resources. In addition, the experiment will show us which if any of our four re-ranking algorithms (see section 5.3.6) provides the best ordering of the search results.

7.2. Hypothesis

In this section we present our hypotheses regarding the results of this experiment. Firstly, we believe that the automatically assessed quality score we offer will improve the ranking offered by Feedback Fruit's prototype search engine. This search engine offers a ranking of search results based solely on topical relevance. Therefore, we hypothesize that the partial quality score we introduce in the search algorithm will offer a semblance of user relevance (especially since we introduce the perceptions of the users through the input of their preferences) thus improving the ranking of the offered results. We also believe that from the four re-ranking algorithms we propose the one that will perform best is option 4 (Mult_Sqr, see section 5.3.6). Based on the complex nature of quality we believe that this complicated branched formula will be better able to deal with the insertion of quality assessments in the ranking algorithm.

7.3. Measurements

Since the purpose for designing this quality model is to facilitate educators of TU Delft in finding high quality educational resources, in this experiment we measure their satisfaction. Specifically, we present the participants with five rankings (the initial ranking offered by Feedback Fruit's search engine and the four re-rankings we propose) of the search results from a query of their choice. Then they are asked to give us their satisfaction with these orderings on a scale from 1 to 5 (where 1 means Least satisfied and 5 Most satisfied). After collecting all these results we do a statistical analysis and find which of the 5 algorithms performed best. The use of user satisfaction as a measure was chosen because one of the most important influencing facets of quality is the fact it is relative. Therefore since quality is intrinsically linked with the satisfaction of the user it is the appropriate metric to determine whether the proposed ranking algorithms improve the existing algorithm that is only based on relevancy.

7.4. SET UP

In this section we describe how we prepared and set up this experiment. First we present the participants of this experiment and then we explain how we prepared before each evaluation.

PARTICIPANTS

For these evaluations we recruited the help of 22 personnel members from TU Delft. Specifically the participants in these evaluations were 4 Teaching Assistants(T.As), 2 Phd students, 13 Assistant Professors and 3 Associate Professors. The participants are specialized in a variety of academic fields, like Computer Science, Aerospace Engineering, Industrial Design engineering etc. Furthermore, the participants come from varying cultural backgrounds. Some are Dutch, Canadian, Mexican, Cuban etc.

SET UP

Before conducting this experiment a digital evaluation form was prepared (here is a link to this form).This form contains nine sections. The first section concerns the collection of general information from the participants. Specifically, the participants were asked to fill in their nationality and profession. This information was asked due to our findings from the literature review concerning the influence of culture and educational level on quality perception.

The following section is designed to collect the scores given by the participants regarding their satisfaction with each of the five ranking algorithms we offer. These five algorithms are: the initial algorithm that orders results based on topical relevancy (Algorithm A) and the 4 re-ranking algorithms (B-E correspond to options Weighted_Sum, in section 5.3.6 respectively) we propose that combine the topical relevancy of each resource with the partial quality score we automatically calculate for it. Specifically, in this section we ask the participants to pick a purpose of search from the two offered ("Prepare a lecture presentation" and "Expand knowledge"), insert their preferences by selecting a distribution of weights of importance for the criteria we use to assess quality and pick a query keyword from a list we provided. Then we present the five different orderings produced by the 5 ranking algorithms and ask them to assign a score from 1 to 5 on their satisfaction with each.

In section 3.4 we explain that direct access to the MIT resources was not possible. Instead we only had access to the metadata of any MIT resources that appeared in a search query. However the available metadata for the MIT repository are seriously lacking (see section 3.4 table 5) so to make the comparison between resources from both repositories (MIT and TU Delft) fair, before each evaluation we created a list of keywords relevant to the field of study of the participant. For each keyword in that list we gathered the links of the MIT resources and preprocessed them to get information regarding their length (number of pages). In addition, in order to provide a score for the criterion source we need to know the field of study a keyword relates to. Since a taxonomy of the terms offered in Feedback Fruit's search engine has not been successfully implemented yet, we had to manually match each keyword offered to the participants to the appropriate field of study. All of this forced us to limit our participants' choice of keywords to a list we had created prior to the evaluation (see appendix C).

The next 6 sections of this evaluation form contain questions regarding the interface and functionalities offered in the prototype tool we implemented. These results however are not pertinent to the goals of this thesis so we will not present them. The final section of this form was an open-ended question about additional remarks where participants were free to offer suggestions or criticisms.

7.5. HOW WAS THE EXPERIMENT CONDUCTED?

The experiments were conducted one on one in a structured manner by following the order of the evaluation form. After a brief explanation of the goals of this project the participants were given access to the prototype tool we implemented and asked to follow the instructions on the evaluation form. During the evaluation participants were asked to select one of the offered purposes of search. Then they had to insert their preferences regarding how quality should be evaluated and pick one search query from a list of possible queries (see appendix C). Using their selections they were asked to perform a search. Then we showed the participants how they could view the different orderings of the search results, namely by pressing 1-5 on the keyboard. They were given free reign on how to use the tool in order to determine their satisfaction with the orderings. Most participants used the titles of the search results and the "Information box" functionality to determine their satisfaction with the various orderings. And some accessed the resources in order to review the material offered. While reviewing the different orderings of the search results, we stressed to the users to keep the selected purpose of search in mind. After going over the results the participants gave a score of their satisfaction on a Likert scale (1:Least satisfied - 5:Most satisfied). Then they proceeded with the rest of the evaluation regarding the interface and functionalities of the offered tool. At the end of the evaluation if the given scores of satisfaction were generally low for all five algorithms the participants were asked to elaborate on their dissatisfaction.

7.6. RESULTS

After the conclusion of the evaluations, we gathered the satisfaction scores of the participants for each ranking algorithm. In figures 7.1, 7.2, 7.3, 7.4 and 7.5 we present all the scores we gathered for each algorithm:

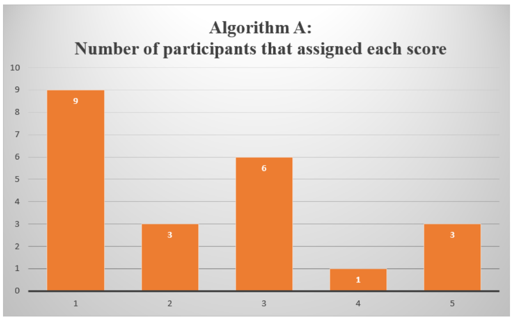


Figure 7.1: Figure that shows the satisfaction scores given to Algorithm A

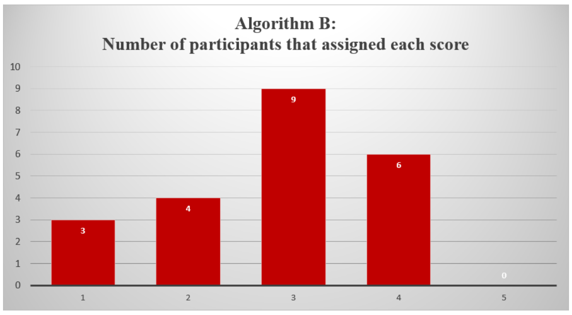


Figure 7.2: Figure that shows the satisfaction scores given to algorithm B

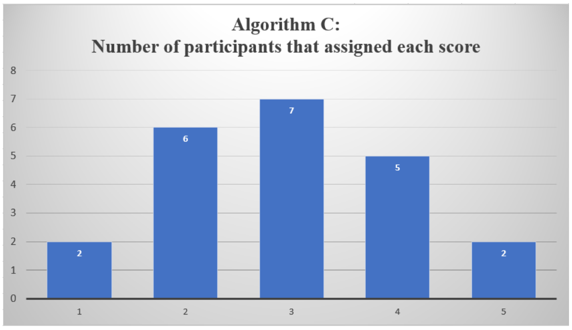


Figure 7.3: Figure that shows the satisfaction scores given to Algorithm C

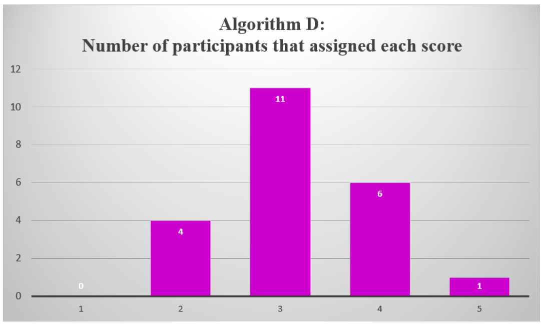


Figure 7.4: Figure that shows the satisfaction scores given to algorithm D

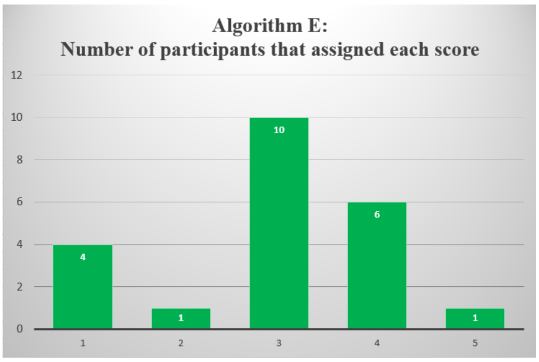


Figure 7.5: Figure that shows the satisfaction scores given to algorithm E

During the evaluations a lot of the participants verbally communicated their satisfaction with the type of results that appeared as a result of our re-ranking algorithms. More precisely, they commented that the appearance of educational materials for the preparation of lecture presentations and articles, books and conference papers for expansion of knowledge at the top of the results was good.

7.7. INTERPRETATION OF RESULTS

In this section we present the analysis of the results we have gathered. Specifically, we do a statistical analysis of the satisfaction measurements of each ranking algorithm. The results of this analysis are shown in table 7.7, specifically this table shows the average score of satisfaction and standard deviation of the measurements given to each ranking algorithm (meaning the initial search algorithm and the four mathematical formulas we propose).

	Algorithm A	Algorithm B	Algorithm C	Algorithm D	Algorithm E
Average Score	2.36	2.82	2.95	3.18	2.95
Standard Deviation	1.43	1	1.13	0.79	1.13

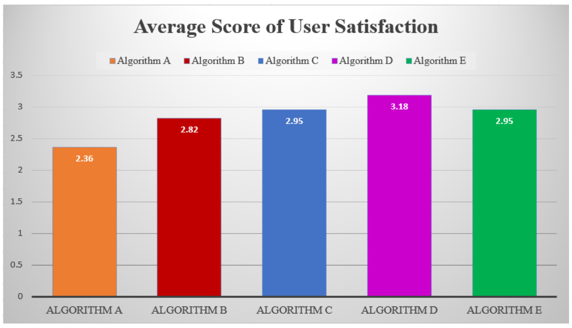


Figure 7.6: Figure that shows the average scores of satisfaction for the five ranking algorithms

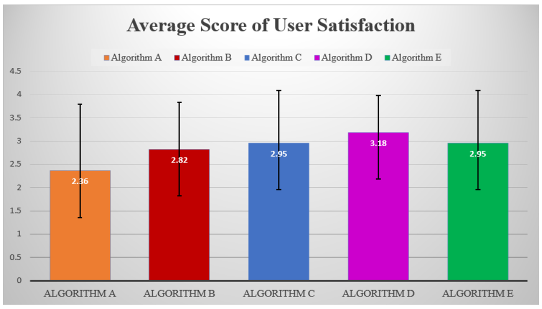


Figure 7.7: Figure that displays the average scores of satisfaction and standard deviation errors of the five ranking algorithms

The results of table 7.7 are displayed in the figures 7.6 and 7.7. Looking at these results we can see that all four re-ranking formulas we propose in section 5.3.6 outperform the initial ranking algorithm used by Feedback Fruits. This, as we hypothesized is due to the fact that algorithm A considers only the topical relevancy. On the contrary our four re-ranking algorithms consider both the topical relevancy offered by algorithm A and the partial quality score we calculate automatically. The fact that all four proposed formulas for the combination of topical relevancy and quality improve the ordering of the search results is evidence that we have succeeded in automatically assessing quality of educational material(answer to research question 2). Although the low average scores of satisfaction suggest that our quality assessment is not ideal and still requires further study.

After the evaluations, when we asked participants who gave low scores to all five algorithms "why?", the majority of them said that the materials resulting from the search query were not relevant. We can think of two reasons why this occurred. Firstly, the fact we limited the participants' choices of query keywords and provided them with keywords that covered a broader field of study. This was unavoidable in some cases, due to the lack of large number of resources. More precisely, in some of the fields like policy management and architecture the material available for more focused, precise keywords were too few making re-ranking futile. Therefore, we had to use more general query keywords. The second reason for this occurrence may be the algorithm used by Feedback Fruits to find relevant resources. Specifically, the use of only topical relevance results in a matching of the keyword with documents that contain this word, however it does not take into account user relevancy. So, although our partial quality score calculation offers some semblance of user relevance in order to fully succeed in offering meaningful re-rankings personalization is needed. For example, use of a person's view or download history, or even previous searches would greatly benefit the search process. One of the participants gave low scores because he failed to use the tool properly. Specifically, the participant said that the only criterion he used to select material was their up-to-dateness however instead of creating his own set of weights giving more importance to creation date, he chose one of the offered set of weights that focuses equally on the source, rating and creation-date criteria.

In regards to the performance of our re-ranking formulas. Between the four algorithms, algorithm D seems to be the best. However the average score of algorithm D, is only slightly better than those of algorithms C and E. So, to ascertain which of the algorithms best combines topical relevancy with quality assessments we look at the standard deviation of the three algorithms. Looking at these calculations we can conclude that algorithm D is indeed the algorithm that best combines relevancy and quality. Although we expected algorithm E to be the one that best combines the two parameters the outcome is logical. From the literature survey we identified relevancy (topical relevancy) as one of the most important criteria for determining quality. Therefore, it is logical that algorithm D which uses a formula that focuses on relevancy (relevancy is squared) performs better in comparison with the other suggested solutions. With algorithm D all resources that have low relevancy are clearly separated from those that have high relevancy. This means that at the top of the resulting list we are left with the high and medium relevancy resources and the partial quality score is used to make distinctions between them. Whereas the other three formulas allow a possible overlap between resources of medium and low relevancy if the partial quality score is high enough. The comments of some users regarding the suitability of the type of resources that appeared when using our re-ranking algorithms suggests that the fitness to purpose parameter is an essential factor in quality quantification. This is logical considering quality's dependency on the purpose of use.

Finally, the results of this experiment answer the final research questions of this thesis. Although from a small sample of our target audience the results prove that we can improve the ranking of the results of some search engines. From this we get positive evidence that our automatic quality assessment is significant enough to improve the ordering of the results (answer to research question 2). This is evidenced by the fact that all four proposed re-rankings performed better than algorithm A. We can also conclude that search engines that only use topical relevancy in their search algorithm can be improved (answer to research question 3). As for search engines that use other factors for their ranking, further research is required which answers our third research question.

7.8. Threats to validity

In this section, we present the threats to validity of this experiment. Firstly, the number of participants used although enough to get an indication of the results is too small to fully represent the target audience. Secondly, the quality model we created is based on the user group analysis. More specifically, the weights we use in the quality come from the user group analysis which means that our quality model's accuracy depends on the perceptions of those we surveyed and that was a very small number of people. In addition, the formula parameters used in some of the re-ranking algorithms (see section 5.3.6) e.g parameters 0.7 and 0.3 of option 1, were selected based on our intuition from the literature findings rather than experiments. One of the facets of quality is educational level, however the absence of appropriate metadata to show who the target audience of the resources is, led to that factor not being considered in our quality model. Additionally, educational level of user also colours their perception but for the purposes of this project we assumed teaching assistants, professors and Phd students are of similar educational level. Time constraints and difficulty in finding large enough number of participants for each group led to this assumption. However, further study with enough participants in each group might provide additional sets of weights that capture the nuances of quality perception influenced by education. In hindsight the value used for the length criterion, meaning the number of pages is not good enough to use for comparison between documents due to the existence of different font sizes, templates, formats etc. A better criterion, would be the number of words used or a combination of both the number of pages and word count. The participation in the experiments was voluntary and therefore unavoidably subject to self-selection biases. The majority of participants are proponents of open education and OERs which means they were positively predisposed towards OER. Finally, most of the participants viewed the ordering of the search results from the re-ranking algorithms in the same order (from 1-5) which introduces bias in their responses.

7.9. Conclusions

From the evaluation experiments we conducted we can conclude that at least to a good degree we have succeeded in automatically assessing the quality of educational resources. This is evidenced by the fact that all four re-ranking algorithms that use this value for the ordering of search results outperformed the ranking algorithm that does not. In addition, we have positive proof that use of this automatic quality assessment can improve at least one search algorithm's (the search algorithm used by Feedback Fruit's search engine) performance. However, in order to have definitive proof of this conclusion more experiments are needed. For one the proposed formulas that performed best (C, D and E) should be optimized and then compared again in another set of experiments with a larger group of participants. Also the quality model should be integrated to a variety of search algorithms, that use different ways to determine relevancy (both topical and user) and check if the automatically assessed quality score improves the ordering for all of them. In general though the positive results we got from the experiments show promise for future endeavours in the use of automatically assessed OERs, not just to improve the discoverability of the high quality ones e.g facilitate the accreditation process.

8

CONCLUSION

In this chapter we offer our conclusions from the design, development and evaluation of the presented quality model. First we indicate some limitations we identified from working on this project. Then we offer our conclusions and finally, we suggest future directions for this work.

8.1. LIMITATIONS

In this section we list a number of limitations and threats to validity we have identified during our work on this project.

Currently the quality of educational material is manually assessed usually by large groups of people. This is a very time-consuming and labour-intensive task considering the growing rate with which new resources are created. So automatically assessing the quality of educational resources is a very important task. Unfortunately what my research suggests and design attempts confirmed is that automatically assessing quality is intrinsically impossible due to its very nature. It is very difficult to create an algorithm that takes into account the wide variety of characteristics needed to fully determine quality. Furthermore, it is impossible to create an algorithm that fully captures the preferences, feelings and thoughts of each person regarding quality. Therefore one major limitation of the presented quality model is its inability to offer a well rounded, fully comprehensive quality assessment of the resources.

Another limitation we were faced with during the development of this quality model is the bad metadata tagging practices practised by educational repositories. The quality assessment offered by our model depends on the existence of metadata that provides information on the criteria used for quality quantification. However looking through three different educational repositories we observed these issues:

1. **Not all resources had the same metadata available**. The fact there is no universal metadata standard for educational resources means not all resources had the same metadata information. This resulted in us spending a large part of this project trying to acquire the missing information to make the comparison between resources from different repositories more fair. Additionally, the use of different meatadata standards depending on the repository meant additional processing of resources so they could be of the same format. This was necessary so they could be processed by Feedback fruit's platform.

2. **Even though the metadata tags existed in the repositories, the metadata was sparse.** Even after processing the resources of some repositories we observed sparsity in the available information. Some resources were missing the necessary metadata while others did not and this created a bias in the reranking process. This in turn may have had a negative impact on the evaluation of the quality model.

3. **The metadata offered are not rich enough to deal fully with the automatic assessment of quality.** This affected the way quality is quantified in the model because we were not able to use all the criteria identified from the literature survey as important. Instead a subset of them is used for quality quantification.

Another limitation of our implementation was the fact we only had access to the resources from MIT during a search query. This meant that some of the processing to get missing metadata like the creation date could only be done during runtime. Additionally, even the non-query, non-purpose dependent criteria were assigned values during runtime. These limitations caused some delays in the calculation of quality scores in our prototype search engine.

During the user group analysis we were faced with some additional limitations. The number of interviews conducted were too few. Although the number of interviews was good enough to get some insight into the thought processes of the users, the majority of the participants all came from the same field of expertise. This means that the qualitative assessment we did regarding their perception of quality (sets of weights offered in the implementation of the quality model) provided a limited view on how educators perceive quality. Furthermore during the interviews it was an oversight not to ask the participants to order the quality criteria separately for each purpose of search. Had we done so, we may have identified more differences in the perception of quality based on purpose besides the change of importance given to copyright licenses .

We also faced another set of limitations during the final evaluation of our quality model and the four mathematical formulas we proposed. Firstly, the limited resources we had access to. Unfortunately, we were only able to get access to educational material from two repositories and this influenced somewhat the evaluation of this model. More specifically, the lack of material in certain fields of study severely limited the list of keywords (see appendix C) that was provided to the participants which in turn caused problems with the evaluation process. We also had only a limited number of participants with which to test user satisfaction of the different rankings of the search results. Additionally, the quality model was created with the limited insights gained from the user group analysis. Therefore the weaknesses of the user group analysis cascaded in the implementation of our quality model affecting its accuracy. Furthermore the parameters and thresholds used in some of the mathematical formulas for quality quantification were decided based on intuition or perceptions from the literature review without scientific proof which also affected the performance of the re-rankings during evaluation. In the quality model we implemented, the factor educational level is not taken into account because of lack of the pertinent information. We also made an assumption that the educational level of professors, Phd students and teaching assistants was of similar level. This assumption is a weakness of the implemented quality model. The use of number of pages as a criterion for length was also not an ideal choice. Participation in this experiment was voluntary and thus subject to self-selection biases. Finally, during the evaluations the order in which algorithms were viewed by participants was the same (specifically the algorithms were viewed in order from A to E) thus introducing another bias.

8.2. Conclusions

In this section we offer our insights from this project and summarize the responses we got for our research questions. First we start with some general insights and then we delve into the results of our research questions.

From our literature review it is obvious that the matter of quality of educational resources has not been properly researched. This is evidenced by the limited literature that exclusively focuses on the matter. In addition, the literature suggests that although quality has been a matter of concern in education for the past two or three decades, it has been used without a full understanding of its meaning and complexity. So, the overall conclusion of our literature survey is that quality is a complex, multi-dimensional concept difficult to define and measure. Also from the literature we concluded that quality should only be defined and considered in relation to a specific context and after careful consideration of the varying parameters (e.g. purpose, context, needs, wishes and expectations of target audience etc.). Trying to define it in abstract terms is infeasible. Furthermore, quality should not be treated in a binary way it is not something that either exists or not but a concept with varying degrees. Basically, quality is determined based on how close a product or service comes to a customer's preconceived notion of excellence. Literature also suggests that there is a need for automated quality assessments of educational material. The ever expanding large-scale repositories that exist do not allow for manual assessment of the content.

From the implementation of our quality model we also came to some conclusions regarding the practical aspect of dealing with quality assessment in an automated way. Specifically, we concluded that most educational repositories follow bad metadata tagging practices. The metadata available in repositories are limited, sparse and do not offer rich information. This affects negatively the automatic quantification of quality we are able to achieve. Also, the lack of a universal metadata standard for educational resources makes their quality assessment from a variety of educational repositories a harder task to implement.

From the research questions of this project, we can conclude the following. The definition of quality and way it is perceived is influenced by numerous factors. Specifically, **quality is dependent on user satisfaction, purpose, context, time, educational level, culture and it is a dynamic concept**. We also offer a well-rounded definition of quality of educational resources taking into account the aforementioned influencing factors. Namely, **quality of educational resources is the dynamic process that shows whether a resource satisfies an ever-changing list of requirements at a specific point in time. Its existence depends on a person's perceptions of whether this resource is relevant and best suits their needs and objectives.** From our research we can also conclude that **fully assessing quality of OERs automatically is an impossible task due to its complex, multi-dimensional nature**. Quality even when put in a specific context depends from so many factors and is so linked to the intricate workings of our brain that we believe **only an approximation of it is possible**. We also identified a list of criteria that can be used to assess quality of OERs. Specifically, we identified the following criteria ***relevancy, content(well-organized, comprehensible, well-presented), reliable, up-to-date, length, number of topics, culture and language, educational level, copyright license,editable, technical characteristics, existence of images and accreditation***.

The comparison between the information we require for a good approximation of a quality score for OERs and the metadata actually available showed there is a mismatch of information. Specifically, the available information is very limited and we require a lot more information in order to improve and optimize the automatic quality quantification we propose (which we think can only be an indicator or approximation at best). Richer metadata are needed. More precisely, access to descriptive, structural, administrative and semantic metadata would allow us to better approximate the full scope of quality of educational resources. From the results of the evaluation of the quality model we implemented it is obvious that **we are on the right track regarding how to combine the quality criteria and quantify quality of educational material**. However the low average satisfaction scores suggest **further research and improvement is needed**. The evaluation also showed that the fitness to purpose parameter is an essential factor in improving the ranking of search results.

From the results of our final experiment we can conclude that **the partial automatic quality assessment we perform has significance and improves the search results of the search algorithm used by Feedback Fruit's prototype search engine**. **More research is needed in the future to test its effect on other search algorithms**. However we can hypothesize that the quality assessment we propose improves the ordering for search algorithms that only calculate topical relevancy. Finally, we suggest that this automatic quality assessment should be used to facilitate some other tasks e.g. the accreditation processes which are mostly achieved manually.

8.3. FUTURE DIRECTION

In this section we offer our suggestions on the direction future work should follow in regards to this research topic. Firstly, our observations regarding the insufficiency and absence of current metadata for educational resources show the importance of metadata. Therefore we recommend that we raise awareness on the importance of having a rich, full set of metadata information and the need for better metadata tagging practices. Additionally, the limitations caused by the use of different metadata standards for educational material depending on the educational repository they originate from, suggests the need for a universal metadata standard for educational resources. By adhering to a specific universal standard, additional processing to format metadata would not be necessary. Besides promoting awareness for these needs regarding metadata, future research should focus on the extraction of the required information. This way quality quantification could be better approximated by using all the criteria we identified from our literature survey as important. Furthermore, research on additional quality criteria could be conducted to offer richer information (e.g. attention metadata to track how a resource is used or information regarding the intended audience for an educational resource etc.) that would help us better approximate a full well-rounded definition of quality. We could also focus on different types of resources like video resources whose quality is influenced by a wide variety of technical characteristics besides resolution.

From the results of the experiment we conducted to evaluate the different re-rankings we proposed, it is apparent that more research is needed. Specifically, the low average scores of satisfaction we got, show that further improvement is needed for quantification of quality. Although our evaluation shows promise for the future further research is required to corroborate our initial findings. This could be achieved with the use of larger amounts of resources that fully cover a variety of topics and experiments involving a larger sample of the target audience. Additionally, integration of the quality model in search engines that use a variety of search algorithms (factoring in user relevance or other parameters) and testing of the resulting performance is also an interesting direction for the future. Furthermore, we propose experimentation and evaluation to optimize the way individual scores are assigned to each criterion and to optimize the mathematical formulas used to quantify quality. These optimizations could also help in minimizing the delays we observed in our implementation which are caused by quality calculations.

As for optimization of the assignment of scores to the quality criteria this could involve the use of NLP methods. In future we could use NLP methods to critically assess the structure and comprehensibility of the resources. This would help better approximate one of the most important quality criteria, the content of a resource which at the moment can only be approximated through rating scores. This would mean a better assignment of scores to the criteria and in consequence more accurate quality quantification. If in the future such methods are successful it would also mean minimizing the reliance of quality assessments on peer-reviews, meaning that less input would be required from the user.

Optimization of the quality model could also be achieved through personalization of the quality model. Since quality is so intrinsically linked to the perception of the user, adjusting the model on each individual's specification could improve the re-ranking of the results. In order to achieve this further analysis of the target audiences should be conducted to better discern the various perceptions that exist on the matter and perhaps discover patterns based on culture, age or profession. These experiments could also involve gathering information for a variety of educational tasks to observe the various differences in perception based on the purpose. This would also allow extension of the prototype we implemented in this project.

From the implementation of our prototype search tool we were also able to identify some functionalities that would be beneficial in facilitating the time-consuming task of finding high quality resources. For one the implementation of interactive visualization features provide users better browsing capabilities. For example the word cloud we implemented could be converted into a concept map that allows users to more easily locate educational material that concern a variety of topics instead of limiting them to a single query. Another feature that could prove useful are recommendation systems, comment sections or forums. A sentiment analysis of their content could provide a better approximation of some of the quality criteria we identified e.g. the structure, organization or presentation of an educational resource.

Finally, the findings and results of this thesis suggest that automatically quantifying the quality of educational resources is possible to a certain extent. However to fully exploit its potential not only to improve current search processes but also for other tasks e.g. facilitating the accreditation process, further research is required.

A

Appendix

Figure A.1 shows the spectrum of possible Creative Commons Licenses ranging from the most open (CC0) to the least open (C) license. These represent the most and least desired value of the copyright criterion respectively.

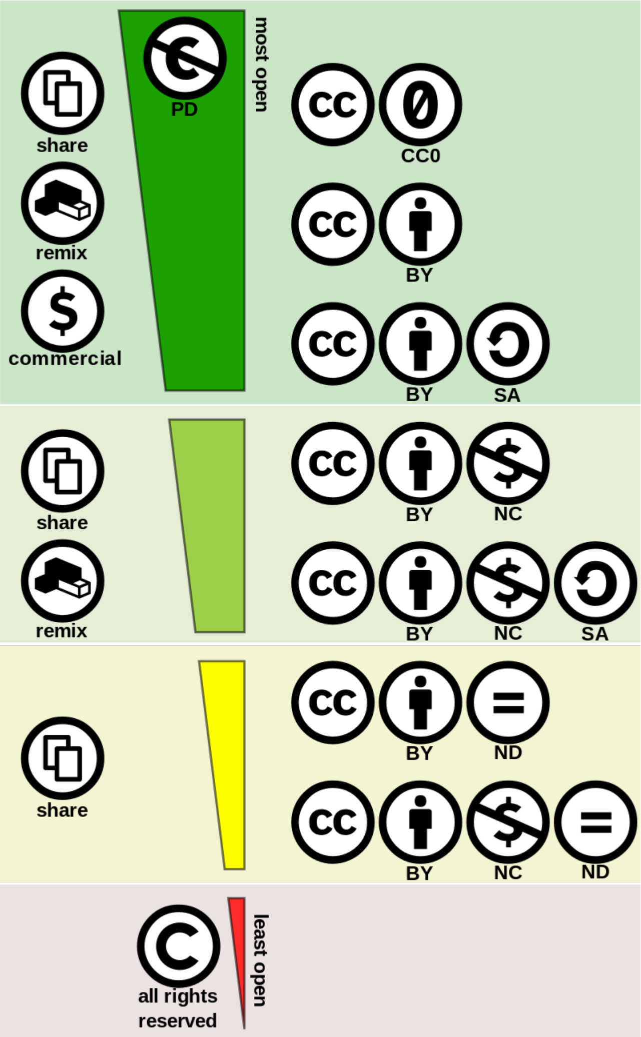


Figure A.1: Creative Commons Licenses spectrum from Most Open to Least Open

Here we provide an explanation of the different components that are used to describe the Creative Commons Licenses:

- "Attribution"(BY): This component means that the individuals and organizations that use the licensed material have to give credit to the original creator of the material.
- "Share-Alike"(SA): The use of this component means that any revised or adapted versions of the material need to be licensed under the same Creative Commons license as the original one.
- "Non-Commercial"(NC): This component means that it is prohibited to use the licensed material for commercial purposes.
- "No Derivatives"(ND): This component denotes that it is prohibited to make any changes to the original material.

B

Appendix

B.1. Questionnaire for interviews

General questions:

1. Do you read the material you find online or do you first download them and then read them more carefully?
2. Do you have difficulty finding high quality educational material? How often do you find the material you are looking for?

Quality Definitions:

1. How would you define quality of educational material? What constitutes high quality educational material to you?

Quality Criteria:

1. Does the issue of copyright licenses factor a lot in your selection of educational resources?
2. Language: When looking for material to expand knowledge do you prefer educational material in your native language or English?
3. Given the following list of criteria: creation date, length, number of topics it relates to, source of material, editable meaning no watermarks or logos, copyright licenses, existing reviews/comments/rating and resolution of images/video. How would you prioritize the criteria regarding their importance for a quality check? If you only had to pick one as the criterion by which you would judge the material which would it be?

Lecture Presentation:

1. What type of materials are you looking for in order to prepare a presentation for a lecture?
2. Do you follow a specific process to collect the material you are looking for or do you prefer to browse until you find something suitable?
3. When looking for educational material do you first look at high quality material then settle for whatever exists or do you change the lecture structure?
4. Do you prefer materials with existing graphs, pictures and tables? Or do you prefer to create those on your own?

Expand Knowledge:

1. What type of materials are you looking for in order to expand your knowledge on a specific topic?
2. Do you prefer materials that combine video and textbook for extending knowledge of students?
3. Do you follow a specific process to collect the material you are looking for?

C

Appendix

C.1. List of available query keywords for evaluation

Mathematics

1. Statistics
2. Circle
3. Polynomial
4. Logarithm
5. Exponential
6. function
7. Derivative
8. Geometry
9. Set theory
10. Integral
11. Differential equation
12. Gradient
13. Graph Theory
14. Calculus
15. Mathematics
16. Function (mathematics)
17. Optimization (mathematics)
18. Matrix (mathematics)
19. Matrices
20. Viterbi algorithm
21. Algorithm
22. Dijkstra's algorithm
23. Linear algebra

24. Linear regression
25. Line segment
26. Probability theory
27. Probability distribution

COMPUTER SCIENCE AND ENGINEERING

1. Computer network
2. Network theory
3. Cryptography
4. Interpolation
5. Raster graphics
6. Game
7. Computer graphics
8. Peer review
9. System dynamics
10. XML
11. Dynamic programming
12. Greedy algorithm
13. Decision making
14. Systems engineering
15. Visualization
16. Metadata
17. Software architecture
18. Pattern recognition
19. State diagram
20. Artificial intelligence
21. Graphic design
22. Decision support system
23. Virtual reality
24. Semantic Web
25. Medical imaging
26. Data visualization
27. Comparison of relational database management systems
28. Database management system
29. Relational database

30. Oracle Database
31. Data analysis
32. Database
33. Data model
34. Data modelling
35. Viterbi algorithm
36. Algorithm
37. Dijkstra's algorithm
38. Linear programming
39. Signal Processing
40. Computer simulation
41. 3D computer graphics
42. Computer network
43. Cloud computing

Aerospace Engineering

1. Aerospace engineering
2. Astrodynamics
3. Aerodynamics
4. Flight dynamics

Physics

1. Magnetic field
2. Magnetism
3. Electronic band structure
4. Quantum field theory
5. Particle physics
6. Quantum mechanics
7. Angular momentum
8. Conservation of energy
9. Fluid dynamics
10. Spectroscopy
11. Magnetic field
12. Optics
13. Infrared spectroscopy
14. Energy storage
15. Electric charge

16. Photovoltaics
17. Electric vehicle
18. Fuel cell
19. Physics
20. Elasticity (physics)
21. Astrodynamics
22. Aerodynamics
23. Drag (physics)
24. Thermodynamics
25. Electromagnetism
26. Waves

Industrial Design

1. Design methods
2. User interface design
3. Industrial design
4. Interaction design
5. Ergonomics
6. Engineering design process
7. Design engineer
8. Sustainable development
9. Aeronautics

Civil Engineering

1. Design management
2. Design
3. Construction
4. Architecture
5. Real estate appraisal
6. Real estate investing
7. Affordable housing
8. Sustainable development
9. Real property
10. Urban planning
11. Urban design
12. Urban studies and planning
13. Building engineering
14. Architectural design
15. Landscape architecture

Biology and Chemistry

1. Condensed matter physics
2. Crystal structure
3. Biology
4. Polymer
5. Molecule
6. Molecular diffusion

Technology, Policy and Management

1. Systems Engineering
2. Infrastructure
3. Decision Making
4. Innovation
5. Ethics
6. System dynamics
7. Peer review
8. Risk management

Bibliography

[1] D. Wiley, *Expert meeting on open educational resources,* Centre for Educational Research and Innovation (2006).

[2] D. Wiley, *The access compromise and the 5th r,* https://opencontent.org/blog/archives/3221.

[3] *The 4r model improved: 5r model of openness,* http://www.e-learn.nl/2014/04/06/5r-model-of-openness.

[4] B. Hegarty, *Attributes of open pedagogy: A model for using open educational resources,* Educational Technology , 3 (2015).

[5] M. Menon and P. Bhandigadi, *Study on cost-efficiency and quality of oer integrated course materials,* (2018).

[6] U.-D. Ehlers, *Extending the territory: From open educational resources to open educational practices,* Journal of Open, Flexible, and Distance Learning **15**, 1 (2011).

[7] D. Wiley, T. Bliss, and M. McEwen, *Open educational resources: A review of the literature,* in *Handbook of research on educational communications and technology* (Springer, 2014) pp. 781–789.

[8] E. Ossiannilsson, K. Williams, A. F. Camilleri, and M. Brown, *Quality models in online and open education around the globe. State of the art and recommendations* (Oslo: International Council for Open and Distance Education, 2015).

[9] W. B. Croft, D. Metzler, and T. Strohman, *Search engines: Information retrieval in practice*, Vol. 520 (Addison-Wesley Reading, 2010).

[10] *Semantic metadata – do you need it, or even want it?* https://www.conceptsearching.com/transitioning-from-semantic-metadata-to-intelligent-metadata/.

[11] N. Elassy, *The concepts of quality, quality assurance and quality enhancement,* Quality Assurance in Education **23**, 250 (2015).

[12] L. Harvey, *Analytic quality glossary, quality research international,* http://www. qualityresearchinternational. com/glossary (2004).

[13] Y. Cheong Cheng, *Quality assurance in education: internal, interface, and future,* Quality Assurance in Education **11**, 202 (2003).

[14] J. Newton, *Views from below: academics coping with quality,* Quality in higher education **8**, 39 (2002).

[15] Y. Cheong Cheng and W. Ming Tam, *Multi-models of quality in education,* Quality assurance in Education **5**, 22 (1997).

[16] E. Sallis, *Total quality management in education* (Routledge, 2014).

[17] N. Seth, S. Deshmukh, and P. Vrat, *Service quality models: a review,* International journal of quality & reliability management **22**, 913 (2005).

[18] E. Barbera, *Quality in virtual education environments,* British Journal of Educational Technology **35**, 13 (2004).

[19] F. Demir, E. Ozsaker, and A. O. Ilce, *The quality and suitability of written educational materials for patients,* Journal of clinical nursing **17**, 259 (2008).

[20] S. Ozkan and R. Koseler, *Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation,* Computers & Education **53**, 1285 (2009).

[21] G. Ghinea and S. Y. Chen, *Perceived quality of multimedia educational content: A cognitive style approach,* Multimedia systems **11**, 271 (2006).

[22] A.-M. Vercoustre and A. McLean, *Reusing educational material for teaching and learning: Current approaches and directions,* International Journal on E-learning **4**, 57 (2005).

[23] E. Duval, *Learnrank: Towards a real quality measure for learning,* in *Handbook on quality and standardisation in E-learning* (Springer, 2006) pp. 457–463.

[24] A. F. Camilleri, U. D. Ehlers, and J. Pawlowski, *State of the art review of quality issues related to open educational resources (OER)* (Luxembourg: Publications Office of the European Union, 2014).

[25] S. Bethard, P. Wetzer, K. Butcher, J. H. Martin, and T. Sumner, *Automatically characterizing resource quality for educational digital libraries,* in *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* (ACM, 2009) pp. 221–230.

[26] M. Custard and T. Sumner, *Using machine learning to support quality judgments,* D-Lib Magazine **11**, 1082 (2005).