Users' Relevance Criteria for Videos in Leisure Contexts

Sarah Ahmed Albassam
Department of Computer and Information Sciences
University of Strathclyde

This dissertation is submitted for the degree of
Doctor of Philosophy

2018

Declaration

'This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.'

Signed:

Date:

Acknowledgements

First and foremost, I am grateful to God who gave me the strength to keep going and complete the PhD. This thesis could not be completed without the support of many people who I wish to thank.

I deeply appreciate my supervisor, Prof. Ian Ruthven for his guidance, patience and encouragements. Despite his incredibly busy schedule as he was the head of the Computer and Information Science department for the majority of my PhD duration, he always made time to meet me when I asked to and provided me with feedback for submitting work. Without his valuable feedback, the completion of this thesis would not be possible. I also appreciate SISGR research group members for their insightful comments during the presentations I gave about my work.

I also want to thank King Saud University for the generous scholarship that enables me to carry on the PhD in the United Kingdom.

My parents, Ahmed and Amal, who always encourage me and who taught me the value of education since I was a child. Thanks for all your support, love and prayers. I would also thank my sisters, brothers and friends, without them I would be very lonely during the PhD journey.

Last but not the least, I would like to thank my husband Bader, who I have shared this journey with. Thanks for your understanding, encouragements and for always being there for me. I would like to thank my sons, Abdulaziz and Faisal for just being them and for their understanding when I am not always there for them.

Abstract

Relevance is a fundamental concept in the Information Retrieval field. As part of the relevance judgment process, users apply several relevance criteria to judge retrieved objects. Relevance judgments and the criteria used to make these judgments are known to be dynamic. Many research attempted to explore relevance criteria users apply when making relevance judgment decisions and examine the dynamic selection of relevance criteria at different stages of the search process. Previous relevance criteria studies focussed on work contexts and the information judged was mainly in text format; with the result that little is known about relevance criteria and its dynamic aspects when applied in leisure contexts, specifically for video content.

The purpose of this research is to understand how typical users of YouTube judge the relevance of videos in leisure contexts; what are the reasons users give when judging video material as relevant or not relevant? Furthermore, the research investigates the dynamic aspects of relevance criteria by examining the differences in relevance criteria at the stages of selecting and viewing videos for leisure.

This research encompasses two main studies. In the first study, a naturalistic diary was performed in which 30 participants completed diaries providing details on their video relevance criteria. In the second study, 24 participants were asked to search YouTube for leisure purposes followed by a semi-structured interview to elicit relevance criteria usage at different stages of the search process.

In total, 28 relevance criteria were identified through the analyses of the diaries' contents and they were grouped into eight categories. The findings revealed that criteria related to the content of the video are the most dominant group of criteria with Topicality being the most dominant criterion. There is a considerable overlap between leisure relevance criteria and previous relevance criteria studies in academic or work-related contexts, but the importance of these criteria varies among different

contexts. New criteria, e.g. Habit, emerged from the data which tend to be more related to leisure contexts. The findings of the dynamic use of relevance criteria study showed significant differences between the selecting and viewing stages in term of the use of relevance criteria with, some criteria being preferred in the selection stage while others are more important at the viewing stage of video interaction. The findings also demonstrated that applying different methods (naturalistic diary study and recorded sessions with interviews) revealed similar and consistent findings.

This research attempted to enrich the current literature by investigating users' video relevance criteria in leisure contexts. Understanding the changes in relevance criteria during the search process provides new insights into the dynamic aspects of relevance judgment and aids the design of information retrieval systems.

Table of Content

List of Tables

List of Figures

List of Publications

Albassam, S. A. A., & Ruthven, I. (2018). Users' relevance criteria for video in leisure contexts. *Journal of Documentation*, *74*(1), 62–79. https://doi.org/10.1108/JD-06-2017-0081

Albassam, S. A. A., & Ruthven, I. (under revision). Dynamic aspects of relevance: differences in users' relevance criteria between selecting and viewing videos during leisure searches, *Journal of Information Research.*

Chapter 1

Introduction

Relevance is a core concept in Information Retrieval (IR). The ultimate goal of all IR systems is to retrieve relevant objects that satisfy the end user's needs (Saracevic, 2016). In Information Science relevance is defined as "a relation between information or information objects (the Ps) on the one hand and contexts, which include cognitive and affective states and situations (information need, intent, topic, problem, task; the Qs) on the other hand, based on some property reflecting a desired manifestation of relevance (topicality, utility, cognitive match; the Rs)" (Saracevic, 2007a). Relevance is considered as a substantial issue at several stages of the information retrieval process, starting from designing and functioning to the evaluation of the IR systems (Borlund, 2003; Ruthven, 2005).

Information retrieval systems performance is evaluated by their abilities to retrieve relevant materials to the user's request. Previously, precision and recall were the well-known metrics used to measure relevance without considering the user. Later on, studies found these measures alone are not sufficient in addressing the problem and advocated to consider the user and his perception towards relevance. Thus, more research which address user relevance appeared (Barry, 1994; Savolainen, 2009; Xu & Chen, 2006). One of the relevance definitions which takes the user into account when defining relevance is "the user's decision to accept or reject information retrieved from an information system" (Schamber 1994, p.3).

Thus, many studies have attempted to understand how people judge the relevancy of retrieved documents resulting in an enormous amount of work concerning user relevance judgement, behaviour and factors affecting their relevance decisions. In the area of user relevance, researchers are especially interested in what reasons users give to judge documents as relevant or not relevant during the relevance judgment process. In other words, what are the relevance criteria that users apply when making a relevance judgement decision? (Saracevic, 2007b)

Previous research (Bateman, 1997; Schamber, Eisenberg, & Nilan, 1990) has shown that relevance is dynamic and that user's relevance judgments can change over time. This evolution in relevance judgment is a reflection of the evolution in relevance criteria choices. The majority of the previous literature investigated relevance criteria in academic or work-related contexts and mainly for textual content. Few studies have attempted to investigate relevance criteria of different media, e.g. image (Choi & Rasmussen, 2002) or video (Yang, 2005) or in everyday life contexts (Xu, 2007).

Therefore, this research attempts to investigate users' video relevance criteria in leisure contexts and the dynamic use of these criteria at different stages of leisure search.

1.1 Problem Statement and Significant of the Research

Although there are rich literature that considered different aspects of relevance criteria, there are still some directions that have not been completely investigated. There is a lack of knowledge regarding how users apply relevance criteria in video relevance judgments. A few studies have focused on this issue, however; these studies were limited to work-related tasks.

Furthermore, the affordability of technology and the Internet facilitate the use of IR systems not just for work-related tasks but also during leisure time (Elsweiler, Wilson, & Lunn, 2011; Hartel, 2003).

This phenomenon has attracted the research community recently and researchers started to identify leisure concept and investigate user's behaviours in leisure contexts. However, few studies have investigated the relevance criteria that participants apply when using IR systems beyond the academic or work-related contexts. Therefore, video relevance criteria in leisure context is an interesting area to be studied.

The importance of this study lies in its attempt to enrich the current literature by investigating users' video relevance criteria in leisure context. In addition, examination of the dynamic aspects of relevance criteria in this context. By shedding light on this uncovered area of research and by examining the dynamic use of relevance criteria, the IR community will gain a deeper understanding of how users make their relevance judgment decisions in leisure contexts, which might be different from work-related contexts. Furthermore, the findings of this study will have an implication on the design of IR systems. Retrieval systems designed for users with leisure needs might be different from those targeting users with work-related needs. The comparisons of relevance criteria mentioned among these different media and situations will guide designers of different IR systems. As leisure users have different needs and motivations, investigating relevance criteria in a leisure context will help in designing systems that address leisure users' needs.

Understanding how relevance criteria selections evolve as the search progress and what criteria are more important at specific stages of the search will provide a deeper understanding of the dynamic aspects of relevance criteria and will have implications on the design of IR systems. IR systems should be more adaptive to the change in users' preferences of relevance criteria during the search and support users with useful information needed for relevance judgment decisions as they progress in their search.

1.2 Research Questions

The main goal of this study is to investigate the criteria users apply in making relevance judgment decisions when searching videos in leisure contexts. This research also aims to investigate the dynamic use of relevance criteria at different search stages. In order to aid in design of more useful IR systems.

To achieve this aim, this study will address the following research questions:

- RQ1: What are the relevance criteria users apply when judging videos in a leisure context?
    - Subsequently, which relevance criteria are most important when judging videos in leisure context?
- RQ2: To what extent do these criteria match the criteria mentioned in the previous literature of text retrieval and/or work task context?
- RQ3: What is the difference in employing relevance criteria between the selecting and viewing stages of video/leisure contexts search?
    - Subsequently, are there significant differences in applying relevance criteria between the selecting and viewing stages of video/leisure contexts search?
- RQ4: Do different research methods provide different or similar findings? Does the diary method provide different findings from recorded search sessions with interviews?

The stages mentioned in RQ3 are explained as follows. The selection stage precedes the actual viewing of the video where participants select videos to watch from

YouTube homepage, specific channel page or from the search result list of their queries. The viewing stage is the actual watching of the video.

1.3 Research Synopsis

Two studies form this research. The first study followed a naturalistic approach and aims to capture users' relevance criteria in the participants' natural setting. Thus, diary was used as the data collection method for the first study. Thirty participants were asked to fill out diaries for a duration of one week indicating their relevance criteria for selecting and watching videos in their leisure time. Additional context-related information was also required such as the topic or motivation of the leisure search and the titles of the selected videos. Relevance criteria were not given to the participants, they were extracted from the diaries. The participants were asked to report on the reasons that make them select the videos to watch and in cases of stop watching a video (which indicate the video is no longer relevant), participants were also asked to report the reasons. So I could have a full picture of the use of relevance criteria including the positive and negative mentions.

The second study followed a more controlled approach, 24 participants were invited separately to a private room in the university and asked to conduct a video/leisure search for 20 minutes. A search scenario was provided to the participants prior to starting their searches. All participants' search sessions are recorded using screen recording software. Interviews are conducted after the searches where participants watch back their sessions and explain their relevance criteria at different stages of the search process.

1.4 Structure of the Thesis

This chapter provided the goal and significance of this research. It showed how little is known about the use of relevance criteria when people search for videos in leisure contexts and how these criteria might change at different search stages. The remainder of the thesis is structured as follow: Chapter 2 reviews the existing

literature on relevance. The chapter begins by discussing the main concepts of relevance and the main directions in user relevance literature, followed by existing studies in relevance criteria and the methods applied in these related studies. Then the chapter discusses the leisure concept and the main forms of leisure. The chapter also presents related studies in leisure and the relationship between leisure and information behaviour studies and the role images and videos play in leisure search. The chapter then concludes by a discussion about how leisure could be related to relevance criteria studies. Chapter 3 introduces the research methods applied in this research. The overall design of this research is mainly based on qualitative methods using different data collections methods. The chapter provides justifications of the research design and reports on the pilot studies preceded the actual studies. Ethical considerations and the quality of the research are also provided. Moving from the methodology chapter, Chapter 4 presents the main data analysis method used in this research and provides details on the coding process and building the coding scheme.

Chapter 5 and Chapter 6 depict the findings of the research. Chapter 5 reports on the findings of investigating relevance criteria for videos in leisure context. The findings are related to research questions RQ1 and RQ2. Chapter 6 reports on the findings of the dynamic use of relevance criteria and it answered research questions RQ3 and RQ4. Chapter 7 provides a discussion of the major findings of Chapter 5 and Chapter 6. Finally, Chapter 8 concludes the thesis by summarising the main findings and presents the contributions this research makes to relevance criteria, in addition, the implications this research has for system design. The chapter also points to directions for future work.

Chapter 2

Literature Review

This chapter provides background information for this research by approaching relevance and leisure concepts. Previous works in both fields are presented. The chapter is divided into two main parts: relevance and leisure. The first part started by presenting the main concepts of relevance (Section 2.1.1) followed by the different directions appeared in user relevance literature (Section 2.1.2). Relevance criteria are the main concern of this research, thus, Section 2.1.3 classifies the works in relevance criteria literature into different directions. Finally, Section 2.1.4 discusses the different methodologies applied in the previous relevance criteria literature. Moving into the second part of this review, Section 2.2 started by introducing leisure concept (Section 2.2.1) followed by its relation with information behaviour studies (Section 2.2.1). Then, Section 2.2.3 discusses the relationship between image and video from one side and leisure from the other side. Finally, the relationship between leisure and relevance criteria studies is presented in Section 2.2.4.

2.1 Relevance

As mentioned earlier, relevance is considered as a main concept in the Information Retrieval field. The main purpose of this section is to provide background information of relevance as a concept and to review the main works in this area.

2.1.1 Concepts

Since its first appearance, relevance has produced a huge disagreement regarding the perception of its concept (Borlund, 2003; Ruthven, 2005; Schamber et al., 1990). In the article “how many relevances in information retrieval?”, Mizzaro (1998) stated that “there are many kinds of relevance, not just one”. In general, there are two main approaches towards relevance: system-oriented and user-oriented approaches. The system-oriented view (algorithmic approach) concentrates on the relation between a user query and the retrieved information. It is based on developing ranking and retrieving algorithms that match the user’s query with documents and retrieve the documents with the highest similarity. On the other hand, the user-oriented view focuses on the relation between the retrieved information and the user’s information needs, as well as on the user and his perception regarding relevance (Maglaughlin & Sonnenwald, 2002; Yang & Marchionini, 2005). System-oriented relevance alone is not sufficient in addressing user information needs (Savolainen, 2009; Xu & Chen, 2006). Relevance cannot be simply approached by traditional measures (such as recall and precision) alone. Other subjective and situational factors have to be considered as they have an effect on relevance judgment process (Barry, 1994). As a result, there has been more interest in user-oriented relevance.

In an attempt to reconsider the definition of relevance concept, Schamber et al. (1990) reviewed the literature and identified some aspects about the nature of relevance. The review resulted in three main conclusions about relevance. First, relevance is a multidimensional cognitive concept, which depends on the users' situations and their perception of the information. Many factors might affect these judgments. Second, they found relevance to be a dynamic concept, which depends on the users' judgment of the relation between retrieved information and their information needs at certain point of time. Finally, relevance was claimed to be a complex but measurable concept if targeted from the user's perspectives.

Several studies emerged as an effort to produce a framework that acts as the foundation of relevance. One of the key frameworks of relevance is the one presented by Mizzaro (1997) which perceive relevance to be a relation between four different entities: system: surrogate/document/information, user: problem/information need/request/query, context: topic/task/ and time. Moshfeghi (2012) argued that Mizzaro's framework neglects the important role of intents or motivations of the user (emotion). A more comprehensive relevance framework is presented by (Saracevic, 2007a). In this work, different kinds of relevance (manifestations) have been identified that demonstrate different types of relations: (1) system or algorithmic relevance, where the relation is between the user's query and the retrieved information objects; (2) topical or subjective relevance, which describe the relation between the subject (topic) of the query and the subject (topic) of the retrieved information. Aboutness as judged by users is the underlying principle in this type of relevance; (3) Cognitive relevance or pertinence which describes the relation between user's knowledge state and the retrieved information, or the relation between user need as seen by the user and the retrieved information; (4) Situational relevance or utility, which is the relationship between the retrieved information and user's task; (5) Affective relevance, expresses the relation between the user's intents, goals, emotions and motivations and the retrieved information (Borlund, 2003; Saracevic, 2007a).

In contrast to Mizzaro, Saracevic's perception regarding relevance acknowledges the role of emotions, intents and motivations of the user and considers affective relevance as a standalone type of relevance. Cosijn and Ingwersen (2000), proposed a modified model of relevance based on Saracevic's manifestations in which affective relevance is not seen as a separate type of relevance but as an underlying principle of all other user-oriented types. Borlund (2003) also agreed that affective relevance is a property of all other types of user relevance and it does not considered as a separate type of relevance. In this work, she presented an overview of the different types of relevance in the context of information retrieval evaluation and concluded

that situational relevance is the most satiable type that acknowledges the multidimensionality and dynamic aspects of relevance.

A classification of the different ways user relevance has been tackled in the literature is presented in the next section.

2.1.2 Directions in User Relevance Literature

Because relevance is considered a core concept in the IR field, several studies have dealt with different aspects of relevance. Part of the studies focused on the dynamic nature of relevance, concerning how relevance judgment behaviour evolves at different points of time. Smithson's (1994) work is an example of such studies. In this work, Smithson studied the information behaviour of 22 Master's students who have to complete an assignment for the requirement of their course during three months (one semester). As the goal of the study is to examine changes in relevance judgments at different search stages, three main stages were identified: initial, final and citing. During the initial stage, the participants have to predict the relevance of a document on a six-point scale from its title. The final stage appeared at the end of the process and the relevance judgment is based on the full documents at this stage. Finally, the citing stage is to check whether the participants have included the retrieved documents in the assignment. The findings showed that 82% of the documents judged as relevant at the initial stage remain relevant at the final stage and that 12% of the initial relevant documents have been cited.

Similar to Smithson, Bruce (1994) has recruited graduate students to search for their own real information needs related to their coursework. Three stages of the search process have identified as: before the user search the information system, during the search and after the full documents are retrieved. The goal was to investigate how user makes relevance judgment and how this judgment changes during the information retrieval interaction (Saracevic, 2007b).

Another set of studies focused on the factors affecting relevance judgment behaviour. For example, Ruthven, Baillie, and Elsweiler (2007) investigated the effects of three different factors on user judgment process. Specifically, the factors are: users' knowledge of a search topic, their interest in the search topic and their confidence in assessing relevance for a topic. Questionnaires were used to gather participants' rating of their knowledge, confidence and interest in the given topic. The findings revealed that all the three factors have an influence on the relevance judgment process. In another work, Bell and Ruthven(2004) examine the effect of another factor "task complexity" on user's search behaviour. Thirty participants were asked to conduct search given three task varying in their complexity. The authors found that participants were able to recognize the different levels of complexities and there was a correlation between task complexity and participants assessment of task completion, in particular, the more complex the task is the less completion task assessment provided by the participants.

Other studies have investigated design factors—such as the structure and amount of information displayed on the result list—that might affect relevance judgment behaviour (Balatsoukas, O'Brien, & Morris, 2010). Another example is the effect of order in presenting retrieved documents on relevance judgment behaviour (Huang & Wang, 2004).

Relevance is known to be subjective. As a result, different users might disagree on document relevance. The agreement and disagreement on relevance judgment among different users garnered attention from researchers in the field long time ago. The reason behind this interest in the inconsistency of human relevance judgments is that user relevance is the underlying principle of the IR evaluation measures (precision and recall). IR test is conducted by comparing what is considered relevance by the system (system relevance) to what actually judged to be relevant by the experts users (Saracevic, 2008). Test collections have been developed to facilitate the improvement in IR systems. Researchers in IR field depend on them in evaluating

their IR system effectiveness. However, as test collections have been developed by collecting humans' relevance judgments and given that human relevance is subjective, many critics have surrounded the appropriateness of such subjective measure for IR effectiveness evaluation. Typically, test collections are developed based on expert judges' assessments. However, as experts are not affordable there was an intent to recruit non-experts judges to build test collections. This substitution of a non-expert who lack the knowledge and experience of experts might have an effect on the robustness of the test collections. Thus, studies emerged to examine the differences between primary and secondary assessors. Moreover, studies occurred to investigate the rate of agreement between the two types of judges in order to discover the influence of using non-expert assessors on the validation of test collections.

Many studies (Al-Harbi & Smucker, 2014; Sormunen, 2002; Wakeling, Halvey, Villa, & Hasler, 2016) addressed the rate of agreement between different judges or group of judges. Al-Harbi and Smucker (2014) found that secondary assessors are not certain in their relevance judgements and the study reports on and categorizes the reasons for these differences between primary and secondary assessors. Wakeling et al. (2016) findings revealed that, although there are differences between primary and secondary relevance judgments behaviours, the agreement between both groups of assessors is high.

Saracevic (2008) reviewed the studies that focused on the inconsistency of human relevance judgments and the effects of this inconsistency on the evaluation of IR systems. In this review, he clarified that the IR research community noticed the inconsistency of human relevance judgment at very early works since Gull (1956) conducted the first IR evaluation test to compare the performance of two different indexing systems. Although the goal of the study was to evaluate the two systems, Gull has noticed the disagreement between two groups of users participated in the study. A group of studies (Bailey, Craswell, Soboroff, Vries, & Yilmaz, 2008; Voorhees,

2000) investigated the effect of inconsistency in relevance judgments on the IR performance measurement.

Finally, many studies have focused on the criteria that users employ when making relevance judgments, as it is a way to gain more understanding of the user's needs and behaviours. The next section will explore relevance criteria literature.

2.1.3 Relevance Criteria Literature

A relevance criterion can be defined as "the parameter or value by which users determine the relevance of a retrieved object at a certain point in time"(Borlund, 2003; Schamber et al., 1990). Several studies in relevance literature have focused on the criteria that users apply when making relevance judgments. Investigating relevance criteria is essential to gain more understanding of the user's judgment behaviour and to aid in designing more useful IR systems. Early research has significantly contributed to the user relevance criteria literature and have identified key criteria and synthesized them into different categories. Schamber (1991) mentioned 22 criteria classified into ten categories gathered from 30 occupational users of weather information. In a similar work, Barry (1994) identified 23 criteria grouped into seven categories. The participants were faculty and student users who had search requests related to their work. Both studies used an interview to conduct their research. Later, Barry and Schamber (1998) combined and compared the results of their previous studies and identified ten criteria in common. The following list presents these criteria with a short description of each:

- Depth/Scope/Specificity

"The extent to which information is in-depth or focused; is specific to the user's needs; has sufficient detail or depth; provides a summary, interpretation, or explanation; provides a sufficient variety or volume".

- Accuracy/Validity

  “The extent to which information is accurate, correct or valid”.

- Clarity

  “The extent to which information is presented in a clear and well-organized manner”.

- Currency

  “The extent to which information is current, recent, timely, up-to-date”.

- Tangibility

  “The extent to which information relates to real, tangible issues; definite, proven information is provided; hard data or actual numbers are provided”.

- Quality of sources

  “The extent to which general standards of quality or specific qualities can be assumed based on the source providing the information; source is reputable, trusted, expert”.

- Accessibility

  “The extent to which some effort is required to obtain information; some cost is required to obtain information”.

- Availability of information/ Sources of information

  “The extent to which information or sources of information are available”.

- Verification

  “The extent to which information is consistent with or supported by other information within the field; the extent to which the user agrees with information presented or the information presented supports the user's point of view”.

- Affectiveness

  "The extent to which the user exhibits an affective or emotional response to information or sources of information; information or sources of information provide the user with pleasure, enjoyment or entertainment" (Barry & Schamber 1998, p. 227).

Based on Barry and Schamber (1998), other studies emerged to discover how to use relevance criteria differently. The remainder of this section will identify main directions and classify the works done in relevance criteria literature. The studies are classified based on their motivation and goals. The directions are: factors affecting relevance criteria, relevance criteria choices in different search stages, image relevance criteria, video relevance criteria and other directions.

2.1.3.1 Factors Affecting Relevance Criteria Choices

Group of studies in the relevance criteria literature have focused on the factors that guide the relevance criteria selection process. Indeed, studying these factors are essential to gain more understanding of the user relevance judgment behaviour through approaching the relevance criteria choices decision.

One of the studies that dealt with the factors affecting relevance criteria choices, is the work done by Wen, Ruthven, and Borlund, (2006). In this work, the authors conducted a pilot study to examine the effect of topic familiarity on sources and relevance criteria selections by users. The participants were 18 postgraduate students and they were asked to engage in two search tasks. The search topic of one of the tasks was familiar to users while the other was not. The participants were provided by a list of 12 criteria (the majority of them inherited from Barry & Schamber) and asked to predict the importance of each criterion to the search task on hand. The authors found that participants used more formal sources when searching a familiar topic. In addition, with the familiar task, participants succeeded in predicting which relevance criteria they were going to apply and the predicted

criteria matched the actual criteria applied. On the other hand, with the unfamiliar task, participants were not able to predict the criteria correctly and found that applying some of the criteria was not straightforward, which is a consequence of the lack of topic familiarity. Furthermore, criteria such as Background experience, Currency and Verification/b (the extent to which user agree with the provided information) seem to be the most important criteria in the familiar topic task. On the other hand, Verification/A (the extent to which information provided consistent with other information) ranked as the most important criteria in the unfamiliar topic task.

Another factor that might influence relevance criteria selection is the task type. In order to examine the effect of this factor on users' relevance criteria choices, Tombros, Ruthven, and Jose, (2005) asked 24 participants to engage in searching sessions using three given simulated work tasks. The first task was "background search" where the participants have to search for general background information of a topic. "Demographics of the internet" was the search topic of this task. The second was decision task where the participants need to make decision-based on the information collected during the search, in this study the participants were asked to decide about the best hi-fi speakers available in their own price range. Finally, the third was many items task, the participants were required to provide a list of items or things. In this study, they were asked to list "interesting things to do over a weekend in the city of Kyoto". The study aims to identify the relevance criteria users apply when making relevance judgment and to identify the most important elements that users focus on during their relevance judgment. In addition, the study examined the effect of different tasks and the stage of the search process on the criteria choices. Various data collection techniques have been used in this study: think-aloud, questionnaires, system logging and informal discussion. Content or Topicality was the most mentioned criteria, however, other criteria such as Layout of the page, Depth/Scope and Authority have also mentioned frequently. The results showed an effect of task types on relevance criteria selections e.g. Scope and Depth was more important in the "decision task".

In another study, Balatsoukas and Ruthven (2012) recruited 24 participants to search the web for real information needs. They integrated an eye-tracking system to associate the relevance criteria applied by users with the visual behaviour during relevance judgment of the search result interface of Google. The main aims were to examine the relationship between the use of relevance criteria and the elements that user fixated on and between the use of relevance criteria and the grade of relevance (relevant, not relevant, and partially relevant). However, the study had also revealed an effect of the ranking order on the use of relevance criteria. More specifically, it found that participants applied more criteria in judging surrogates that appeared on Google's first results page. Another finding is that users spent more effort (time and fixation) on non-relevant documents.

To sum up, various factors have been investigated in order to examine their influence on users' relevance criteria selections. Topic familiarity, searching task type and the ranking order of the result list found to have an effect on the users' relevance criteria choices.

2.1.3.2 Relevance criteria choices in different search stages

Relevance is known to be dynamic (Schamber et al., 1990). The same user may find a document relevant to his or her information needs at an early stage of the search process and later, as the search progresses, might judge the same document as non-relevant. Changes in relevance criteria applied in the process of relevance judgment are the reasons behind this dynamic relevance judgment decisions. Many studies have investigated the change in criteria selection and importance among different search stages.

A group of these studies have investigated the dynamic application of relevance criteria for students or academics while they performing assignments or academic-related searches. In her study, Bateman (1997) examined the change in relevance criteria of 35 graduate students while they seek information for a given research paper as part of their course assignment. The participants were asked to rank the

importance of 40 given relevance criteria among Kuhlthau's six stages of information seeking. In contrast to other studies of dynamic nature of relevance, Bateman reported no change in criteria choices among different search stages. This might be due to the limited number of participants (Taylor, Zhang, & Amadio, 2009) and the fact that the study focused only on highly relevant documents.

In the same academic context, Vakkari and Hakala (2000) studied 11 students who were preparing a proposal for a master's thesis and asked them to conduct a search three times at different phases of the project: at the beginning, middle and final phases and they judged both lists of references and full-text documents. A variety of methods and data collection techniques were applied in this study, e.g. interviews, think aloud, searching log files and searching diaries. The study identified six major categories of relevance criteria with 25 sub-categories. The findings showed that relevance criteria for references were more stable than relevance criteria for judging full-text documents among the different stages. In addition, for judging references, the study found Topicality is the most important criterion at all stages of the search process. Recency experienced a decrease in its mentions as the search progress while Interest in specific sources increased. For the full-text documents, Topicality was the most important criterion among all the stages while Personal Interest decreased as the search progress.

Wang and White (1999) conducted a follow-up study as part of a research project which aims to understand the decision making process underlying academics' selection of documents during a research project. The study focused on the change in relevance criteria as the academics progress in the research project, specifically between these three stages of the search process: selecting, reading and citing. Fifteen out of 25 participants who participated in the first study were interviewed. Topicality was the most dominant criterion among all the stages. Some criteria such as Novelty, Expected Quality, and Availability were only mentioned in the reading stage, while other criteria such as Classic/Founder, Standard Reference, and

Credential were only mentioned in the citing stage. The emergence of new criteria in this follow-up study in the reading and citing stages but not in the selection are due to the lack of stimulating information on the bibliography.

Tang and Solomon (2001) conducted laboratory and naturalistic studies to investigate the change in relevance criteria between two search stages: citation and full text. In the laboratory study, 90 undergraduate students were given an assignment and were asked to rank the importance of 15 given criteria when they select documents to use in their assignment at both stages: citation and full text. In the naturalistic study, nine PhD students who were performing a literature search to support their research topics were asked to think aloud while judging the relevance of retrieved bibliography and to complete a form when they judge the full document. The findings from both studies revealed an increase of mentions of the following criteria moving from citation to full-text stages: Importance, Newness (which means Novelty), and Topical Focus criteria, while Recency experienced a decrease in its mentions.

Similar to the previous works, Taylor et al. (2009) recruited 39 participants to conduct searches in a laboratory in order to discover the criteria that are more important in eight different search stages adopted from Kuhlthau and Ellis studies (Ellis & Haugan, 1997; Kuhlthau, 1993). The search task was the same for all participants. Participants were asked to rate the importance of ten predefined criteria to their relevance judgment at these search stages. The study reported on a set of criteria that participants prefer to employ in all stages of the search process. They also reported that some criteria, e.g., Most Recent, gain more importance as the search progresses.

Recently, Cole, Beheshti, and Abuhimed (2017) explored how relevance evolved as students progress through Kuhlthau's stages. The participants were 60 grade eight students who conducted searches related to history class assignments for a duration of four months. The study source of data where the students' assignments themselves. Instead of focusing on relevance criteria to judge retrieved content, the study reported on how the participants constructed knowledge through search

phases based on evolution in topical and psychological relevance. Study findings indicate that psychological relevance is constructed by the students in Associating, Translating and Verticalizing phases. Topical relevance only occurs in the Verticalizing third phase.

Few studies have examined the dynamic aspects of relevance criteria for non-academic tasks such as self-generated search tasks. Savolainen and Kari (2006) and Xie and Benoit (2013) investigated relevance criteria applied by participants when judging links (result list) and web pages (full document) during a web search. In Savolainen and Kari study, nine participants were asked to think aloud while they searching the web for a self-chosen topic. The study found 18 criteria used to judge both links and web pages and those criteria were similar for both stages. The most used criteria were criteria related to the content of the links and web pages. In specific, Topicality and Specificity were the highly used criteria in both stages. The study did not report on how a relevance criteria change as the search progress from one link or page to another. Similar to Savolainen and Kari (2006), Xie and Benoit, (2013) conducted a comparative study regarding user relevance criteria in judging search result list and full documents. Thirty-one general users were asked to complete a questionnaire and to think aloud their thoughts during searching the web for the self-generating task. The results showed that many of the evaluation criteria selected for the list and document evaluation are similar. These criteria include: Scope, Specificity, Reputation, Depth, Credibility, Cost, and Language. The study also reported on a set of criteria that are uniquely mentioned in list or documents. For example, some of the criteria exclusively mentioned in document evaluation include: Unique Information, Currency, Accuracy, Availability, Length and Type. Organization (or rank order) was only mentioned in list evaluation.

Another study which investigated relevance criteria for non-academic tasks is the work of Tombros et al. (2005). The study investigated relevance criteria applied by 24 participants when searching web pages for three different information seeking tasks.

The authors attempted to examine the effect of the type of the search task and the stage of the search process on the criteria choices. In order to accomplish the latter goal, two stages of the search process were identified for each task: the first and last set of web documents visited by the participants. The findings showed a variation in criteria mentions between these two stages and that this variation depends on the task type. For example, at the second search task (decision task), more mentions of Scope/Depth and Page Layout reported at the end of the task.

Another group of studies investigated relevance criteria at different stages of image searches. For example, Choi and Rasmussen (2002) interviewed 38 faculty and graduate students to discover the relevance criteria that they employ when making relevance judgments of images. The authors have also investigated the change in relevance criteria that users employ at two different stages of the search process: before the participants having the results of the search and after. Their findings indicated differences in the priority of applying the criteria with the importance of criteria such as Appeal of Information and Accessibility increased as the search process progress whilst the importance of Topicality decreases.

Hirsh (1999) and Reuter (2007) both examined relevance criteria applied by young people when judging photos. In Hirsh's study, the participants were ten fifth grade students who searched for different resources to find information about a self-chosen sports celebrity in respect to a four weeks project related to a class assignment. In Reuter's study, the participants were 96 elementary school students working in a pair to select books for recreational reading from a digital library. Hirsh interviewed the participants twice at two stages of the search process: early stage of the research process (week 1) and at the end of the research process (week 3). In addition, an observation was held while the participants were asked to think aloud their thoughts during their searching. Reuter observed the students when they were working in a pair to select a book to read from the digital library. Relevance criteria applied in three stages of the search were considered: selecting (result list), judging

(surrogate) and sampling (full text). Think aloud and log data were collected as well. Hirsh found changes in relevance criteria between the two stages. In the first stage, Topicality was the dominant criterion, mentioned by most of the students. Approaching the end of the search process, Topicality decreased and the students mostly mentioned the Interesting criterion. This change in relevance criteria indicated a change in students' knowledge about their sports celebrity as they progress in the search process. Reuter (2007) reported on differences in applying relevance criteria among the stages of the search process. The findings showed that Novelty had high mentions in the selecting stage while Accessibility peaked in judging stage.

One study, Pian, Khoo, & Chang, (2016) identified relevance criteria people apply when searching information in health discussion forums for different purposes: searching information for their own health issue, searching for other people's health issue, and when browsing without a particular health issue in mind. Fifty-eight participants' eye movements were captured at two stages of the search: result list and the full post content. The findings reported on differences in relevance criteria among the three search purposes at both stages of the search but not on differences between the two stages. Table 2.1 summarizes the dynamic relevance criteria studies presented in this section.

Study	**Participant**	**Method**	**Search stages**	**Task**	**Findings**
Bateman (1997)	35 students	Questionnaire	Kuhlthau's stages	Academic	No change in criteria choices among different search stages
Vakkari & Hakala (2000)	11 students	Interviews, think aloud, log file, diary	Beginning, middle and final phases of the project for both references and full-text documents	Academic	Criteria for references more stable than the full text. Criteria for references: Topicality most common among all stages. Recency more common in the beginning while interest is more common in the final stage. For full text, topicality dominant among all the stages while personal interest is more common in the beginning
Wang & White (1999)	15 academics	Interview	Selecting, reading and citing	Academic	Topicality dominant among all stages. Novelty, expected quality and availability were only

					mentioned in reading stage. Founder, credential in the citing stage
Tang & Solomon (2001)	90 students	Questionnaire, think aloud	Citation and full text	Academic	Importance, Newness (which means novelty), and Topical Focus criteria more common in full text. Recency is more common in citation stage
Taylor et al. (2009)	39 students	Questionnaire	Eight stages adopted from kuhlthau and Ellis	Academic	Recency is more common in later stages
Cole et al. (2017)	60 children students	Assignments	Kuhlthau's stages	Academic	Psychological relevance is constructed by the students in Associating, Translating and Verticalizing phases. Topical relevance only occurs in the Verticalizing third phase
Savolainen & Kari (2006)	9 general users	Think aloud	Links and web pages	Self-chosen topic	Topicality and specificity were the most common criteria at both stages
Xie & Benoit (2013)	31 general users	Questionnaire, think aloud	Search result list and full document	Self-generated task	Unique information, currency, accuracy, availability, length and type criteria were only mentioned in the document evaluation. Organization (rank order) only mentioned in list evaluation
Tombros et al. (2005)	24 students	Think aloud	Beginning and end of the search	Pre-determined search tasks	Scope/depth and page layout more common at the end of the search for the decision task
Choi & Rasmussen (2002)	38 faculty and students	Pre and post questionnaire interview	Before and after having the search result	Search for images for professional use	Topicality more common before searching. Appeal of information and accessibility more common after the search conducted
Hirsh(1999)	10 children students	Interview	Early stage (week 1 of the project) End-stage (week 3)	Search for celebrity photo in academic context	Topicality was dominant at the early stage while interesting is more common at the end stage
Reuter (2007)	96 children	Observation, think aloud, log	Selecting, judging and sampling	Leisure reading	Novelty is more common in the selecting stage while accessibility is more common in the judging stage
Pian et al. (2016)	58 general users	Eye tracking	Result list and full post	Search for health information	Findings report on differences between different tasks but not on the differences in relevance criteria between the two stages

Table 2.1 Dynamic relevance criteria studies

In conclusion, this section has shown that many studies have investigated the dynamic use of relevance criteria among different search stages. The findings of these studies reported differences in applying relevance criteria among the search stages with some criteria being more important at specific stages. Despite the number of studies of dynamic aspects of relevance criteria, the majority of these studies were longitudinal and the search task was for students conducting assignments or research project. Furthermore, the vast body of dynamic use of relevance criteria literature was on searching for textual information rather than audio or visual information.

2.1.3.3 Image relevance criteria

Numerous studies have investigated relevance criteria applied by users when making relevance judgment decisions. These studies revealed a list of relevance criteria in text retrieval systems and encouraged researchers to examine whether the same criteria are applied in non-text retrieval systems. Studies have emerged to investigate user relevance criteria in different media, e.g., image retrieval, music retrieval (Laplante, 2010) and video retrieval.

As mentioned in Section 2.1.3.2, Choi and Rasmussen (2002) attempted to discover the relevance criteria that users employ when making relevance judgments of images. They have also investigated the evolving in the selection of the relevance criteria at different stages of the search process. The authors interviewed 38 faculty and graduate students of American history to determine the reasons for their searches and the information they are looking for when judging images from the Library of Congress Memory photo archive which is available for public on the web. Participants were also required to complete questionnaires to indicate the importance of the relevance criteria applied at two points of the search. The findings showed that Topicality dominates among other criteria. In addition, other criteria like Image Quality and Clarity seems to be important to users.

As mentioned before, relevance criteria have been investigated among different demographic groups and occupations. One example of studying image relevance criteria in a certain profession is the work of Markkula and Sormunen (1998). They attempted to understand the searching behaviours of journalists while seeking photos to be used in their real work situations. In particular, relevance criteria applied by the journalists were identified. The authors interviewed and observed eight journalists in order to investigate the relevance criteria that they based their relevance judgment on when searching in a digital newspaper photo archive. The study was conducted at Aamulehti, which is a popular newspaper in Finland. The results showed consistent findings with the previous literature. Topicality found to be the first criterion applied by the participants followed by other criteria regarding the technical and contextual aspects of the retrieved images e.g. the Quality of the Image. Finally, visual attributes such as aesthetic attributes or emotional feelings were investigated at the last stages of relevance judgment process. The findings also revealed and confirmed the situational nature of relevance. In particular, the choice of relevance criteria and the priority in applying them strongly depend on the work situation.

Similar to Markkula and Sormunen (1998) and based on their findings, Westman and Oittinen (2006) studied image searching behaviours for the same specific workgroup (journalists). One of the goals of this work was to identify the importance of different image relevance criteria. Nine journalists were given 47 criteria (extracted from the literature) and asked to rate the effect of each criterion on their relevance judgment decision based on five points scale. It has been found that contextual factors, e.g. the nature of the article, space reserved for the image and publishing section, have influenced the relevance judgment decisions. Moreover, Topicality applied at early stages of the judgment process but its importance decreased as the judgment process progressed. At the final stages, participants acknowledge more personal criteria such as Interest and Fun.

Hung, Zoeller, and Lyon, (2005) have also focused on image relevance criteria of undergraduate students from the journalism and media studies department. However, the authors focused on examining the influence of different search tasks on the relevance judgment decisions. In this pilot study, ten participants were given three image search tasks: find general non-unique photo, specific unique photo and subjective emotional photos. Participants were asked to save images relevant to the tasks on hands. The search was followed by interviews where the participants clarify the criteria that they based their images selections on. Typicality, Emotion and Aesthetic were the most common criteria among the three search tasks. The results showed that participants applied more personal feeling and textual information of the images in the general and subjective tasks while relying on specific image features in the specific image search task.

Similar to Markkula and Sormunen (1998), Sedghi, Sanderson, and Clough (2012) investigated relevance criteria applied by another professional group. The authors interviewed 29 health care professionals in order to extract relevance criteria applied by them when searching medical images. In addition to the interview, the participants were asked to conduct a real medical image search and to think aloud their thoughts while searching. The findings revealed 15 criteria grouped into three categories: visual, textual and other criteria. Similar to the previous studies, Sedghi also found Topicality the most important criterion. Furthermore, Image Quality and Dimensional Size of the image criteria where mentioned frequently by participants. Finally, this study has emphasized the concept of situational relevance, by stating that users apply different criteria in different situations when making relevance judgments.

As mentioned in Section 2.1.3.2, Hirsh (1999) explored relevance criteria applied by children when judging the relevance of the retrieved texts and photos. Besides investigating the dynamic evolution of relevance criteria, the study revealed an interesting finding that might distinguish young people search behaviour. Children

pay less attention to Authority as a criterion in selecting images and have frequently mentioned Peer- interest and Accessibility.

In another work, Hamid and Thom (2010) conducted an exploratory user study which aims to identify relevance criteria applied by users while searching images on the World Wide Web. Twelve participants volunteered to participate in image searching using Google image. Three different image tasks were given to the participants: specific, general and abstract in order to examine any effects of task types on the relevance criteria choices. The participants were also provided by a list of ten predefined criteria (Topicality, Accuracy, Suggestiveness, Appeal of information/ Interest, Completeness, Technical attributes, Emotion, Consequence, Composition/ Strong Visual Impact). These criteria were chosen from the previous literature of relevance criteria. Questionnaires were used to ask the participants to identify which of these criteria were useful in making image relevance judgments. Besides questionnaires, screen recording was also used to gather data. Findings showed that Topicality, Appeal of Information and Composition are the most mentioned criteria among the various tasks, however, participants gave different weights to these criteria among the different task types. Table 2.2 summarizes the image relevance criteria studies mentioned in this section.

Study	Participant	Method	Study Context	Findings
Choi & Rasmussen (2002)	38 faculty and students	Interviews and questionnaires	Searching Library of Congress memory photo archive	Topicality is the most used criterion. Image quality and clarity were also important criteria to judge photos
Markkula & Sormunen (1998)	8 journalists	Interviews and observations	Searching Digital newspaper archive	Topicality is the most important criteria followed by the image quality
Westman & Oittinen (2006)	9 journalists	Questionnaires, interviews and observations	newspaper editorial office	Topicality is more important at early stages while interest and fun are more important at final stages
Hung et al. (2005)	10 students of journalism and media studies	Search tasks followed by interviews	Searching AccuNet/AP Photo Archive	Typicality, emotion and aesthetic were the most common criteria
Sedghi et al. (2012)	29 health care professionals	Interviews and think aloud	General medical image search	Topicality is the most important criterion. Image quality and dimensional size were also important criteria
Hirsh (1999)	10 fifth-grade students	Interviews, observations and think aloud	In the school library, searching for images using online catalogue, an electronic encyclopaedia, an electronic magazine index, and the World Wide Web	Peer-interest and accessibility were common. Topicality is common at early stages of the search
Hamid & Thom (2010)	12 students	Questionnaires and screen recordings	General image search using Google image	Topicality, the appeal of information and composition are the most common criteria

Table 2.2 Image relevance criteria studies

In conclusion, investigating image relevance criteria have shown that many of the predefined text relevance criteria (such as Topicality, Clarity, Quality, Affectivness and Authority) are also applicable to images. Topicality remains the dominant criterion among others, however, its importance decrease as the search progress. In addition to Topicality, Affectiveness criterion seems to have an essential role in

judging images. Moreover, studies of image relevance criteria have contributed in discovering new criteria such as Aesthetic attributes and the Size of the images. Other criteria might be specific to journalists such as the space reserved for the image and the nature of the article.

2.1.3.4 Video relevance criteria

The previous section presented relevance criteria studies in image retrieval. The aim of this section is to shed the light on the limited studies which approached video relevance criteria. One study in the literature which addressed relevance criteria in the domain of video retrieval is the PhD thesis of Yang (2005). The study explored the criteria people applied when making video relevance judgments. In this study, 26 users were participating in three different user tasks: illustration, collection and production tasks. The participants who represent the illustration task were ten professors who need videos to be presented to students in their lectures. The collection task was represented by eight librarians who need to augment the library video collection based on some requests or just to make the collection more comprehensive. The last task is the production which represented by eight video editors who need to produce stories. Yang interviewed each participant in order to identify the relevance criteria they use. More specifically, participants were asked to recall work-related search tasks in order to discover their searching strategies. An optional real search task was also offered where participants think loudly their thoughts and how they reached their relevance decision. Besides identifying user relevance criteria, the study also aims to examine the effect of the different task types on the choices of the relevance criteria. The findings revealed 36 relevance criteria classified into three categories: textual, visual and implicit. For video relevance judgments, Topicality remains the most important criterion. The study also revealed a strong influence of task type on relevance criteria selection. Participants showed different evaluation of the importance of the criteria among the different task groups. For instance, production group have focused more on audio-visual criteria than the two other groups.

Another attempt to discover video relevance criteria is the work of Ju and Albertson (2014). Besides identifying the main criteria, the study aims to discover the ranked priority of each criterion. Free word association method was used to collect the data. Free word association is a method used to collect people's thoughts or reactions regarding a certain concept. Participants have to be prompted by a question regarding the targeted concept and asked to express freely what thoughts came to their minds in relation to this concept. Fifty-two journalism and political science college majors were presented with the following question: "In the context of using an online video website or video digital library, what features and qualities of the website are most important for you to use?" The participants have to answer by giving five words or phrases. Content analysis was applied to the collected data and 28 criteria classified under five main categories were identified. Based on the frequencies of criteria mentioned in the study, results revealed more emphasize on criteria related to Retrieval category followed by criteria related to the User Interface, Collection Quality, User Support, and Organization of Collection respectively. The study focused on participants' perceptions of the properties of the digital libraries or video websites in general but not on how the participants select and judge videos specifically.

Cunningham and Nichols (2008) explored the everyday video information seeking behaviours. In this study, 234 participants were observed and interviewed by undergraduate students of a course on Human-Computer Interaction. The study mentioned briefly some cues or attributes of the video retrieval system that are used in relevance judgment decisions such as Comments, Thumbnails and Number of Views. However, it says nothing about other aspects of personal criteria such as Novelty and Authority.

Chung and Yoon (2012) interviewed and observed 20 Korean college students to investigate their needs and searching behaviours when searching multimedia on the web. The study concern is multimedia search in general and the findings report on the participants' needs, multimedia uses, searching sources, search barriers and

relevance. The study found that relevance criteria that are important to participants when searching multimedia for entertainment are: Topicality, First Ranked Result, File Quality, and Popularity.

Another work which investigates the motivations for searching videos online is the work of Lux, Lagger, and Marques (2011). The aim of this study is to explore the motivations for using videos' websites and how these sites are used. Twenty-two participants were interviewed and asked about their video searching experiences and their motivations for video searching and viewing. The findings revealed that YouTube is the most used website for videos followed by Facebook and that most participants watched videos for leisure (only 20% search videos for professional use). The findings also revealed a link between the triggers or motivation that make people search for videos with the genre of the retrieved video. A summary of the studies presented in this section is provided in Table 2.3

Study	**Study goal and context**	**Participant**	**Method**	**Findings**
Yang (2005)	Video relevance criteria for Work-related tasks	26 academics, librarians and video editors	Interviews with optional real search	Topicality is the most dominant criterion Cinematography and scene-level identified as video relevance criteria
Ju & Alberston (2014)	Evaluate the features of video digital library(C-SPAN)	52 journalism and political science students	Questionnaire (using free word association method)	Retrieval category criteria (e.g., filtering, effectiveness) are the most mentioned criteria. In the second position is the user interface category (e.g. layout, surrogate)
Cunningham & Nichols (2008)	Locate and view videos in the everyday life	20 students	Interviews and observations	Topicality, first ranked result, quality and popularity are the main criteria
Lux et al. (2011)	Retrieve and watch videos online	22 general users	interviews	YouTube is the most used video retrieval system Majority of video search is for leisure purposes

Table 2.3 Video relevance criteria studies

In conclusion, this section has covered the limited studies found about video information seeking and relevance criteria. It shows how video relevance criteria were under-covered by the literature. The majority of these studies which considered video searching investigated the needs and motivations for entertainment video search with limited mentions of the relevance criteria applied in the relevance judgment process. Yang's study considered relevance video relevance criteria in depth, however, the study focused on work-related context.

2.1.3.5 Other direction in relevance criteria literature

The previous sections have investigated relevance criteria works in different directions such as: the factors affecting relevance criteria choices, how the relevance criteria selection might change at different search stages, image relevance criteria and video relevance criteria.

There are some studies which follow different directions. For example, a set of studies attempted to compare the set of relevance criteria used in different situations. Maglaughlin and Sonnenwald (2002) investigated the criteria used in partially relevant judgment and compare them to those used in relevant and not relevant judgments. Basically, users judge retrieved documents as relevant if they succeed in meeting their own needs and non-relevant if they do not. Partially relevant documents are documents in between that neither fully meet user's needs nor failed to satisfy these needs. Twelve graduate students were asked to participate and complete a questionnaire followed by an interview. Twenty documents based on the participants' real information needs were collected and participants judge these documents. The study mentioned 29 criteria grouped into six categories (abstract, author, content, full text, journal/publisher, personal). The findings revealed that partially relevance documents are judged with the same criteria as relevant and not relevant documents and Content followed by Full text are the dominant criteria.

Furthermore, Xu and Chen (2006) applied a different methodology in order to investigate relevance criteria. They examined the validity of five hypotheses each claims a contribution of one of the following criteria to relevance judgment decision (Topicality, Novelty, Reliability, Understandability, Scope). Participants were 130 undergraduate and graduate students and they were provided by four search topics to choose from. They also have the option to create their own topic. The participants chose two documents from the results of the search and evaluate their relevance based on a given survey. The study concludes that Topicality and Novelty were the most criteria mentioned by participants, followed by Reliability and Understandability. However, Scope had no significant contribution to relevance judgment.

In another study, Ju and Gluck (2011) asked 244 users to complete a questionnaire regarding their relevance judgment criteria. The goal was to distinguish core from peripheral criteria. The author reported 26 categories for relevance criteria some of them are core criteria e.g. Pertinence, Reliability, User need and Importance. From the 26 criteria presented in this study, some exactly matched the previous list of criteria provided by Barry and Schamber (1998). For instance, Easiness/Clarity, Accuracy, Currency, Availability and Accessibility. Other criteria mentioned with a different name but still could be mapped back to Barry and Schamber (1998). For example, Completeness/Detailness criterion could be mapped to Depth/Scope/Specificity; Reliability/Credibility could be mapped to Quality of Sources.

To sum up, despite the variety of directions studies mentioned in this section pursued, the resulting criteria from these studies enriched the literature. Some of these studies follow a comparative approach by comparing the criteria among different situations, others approaching the relevance criteria from a different angle by applying methodologies different than the previous literature. The next section will list and compare different methodologies applied in relevance literature.

2.1.4 Methods Applied in Relevance Literature

Both quantitative and qualitative methodologies have been used in the previous studies. Different research methods and data collection techniques have been applied. The majority of the previous relevance research conducted user studies. Another set of studies asked the users to recall a situation when they make a relevance judgment decision without conducting a real searching session. Furthermore, there are studies which follow an experimental approach by setting hypothesis and design an experiment to test them. In all these different approaches, a variety of data collection techniques (e.g. interview, talk aloud protocol, eye-tracking, video recording, questionnaire) were used. The following subsections will elaborate more on data collection methods applied and the different ways of conducting relevance studies.

2.1.4.1 Data Collection Techniques

As mentioned before, relevance literature has applied various data collection methods. Interview, questionnaire, talk aloud protocol, eye-tracking and video recording have been all used to collect data. This section describes briefly some of these methods.

Interview

Interviews are one of the data collection methods that have been used in relevance studies. They are suitable in the case of complex questions and when the researcher would like to gain in-depth answers from participants. There are many types of interviews varying from very structured type where the researcher taking hold of the entire situation, to informal un-structured interviews where the researcher has less control on the progress of the interview and the participant is leading the conversation. It is the goal of the study which directs the choice of the interview category (Pickard, 2013).

Questionnaire

Questionnaires are suitable in the case of collecting data from a large sample as it considered to be a low-cost method. One of its limitations is the lack of direct contact with the participants and as a result, the participants could not clarify any single question and could not seek the researcher's help, in contrast, to interview where the direct interaction with participants exist (Pickard, 2013).

Think Aloud Protocol

Think aloud protocol is a type of data collection techniques where participants are asked to perform a task and verbalize what comes to their minds at the same time. The only investigator (researcher) role is to encourage the participants to keep think aloud their thoughts if they pose for a while. Participant's talk is audiotaped and then transcribed. From analysing transcripts of this verbal content, researchers could identify the information that participants concentrate on in accomplishing the task. The method is applicable to extract rich in-depth data from a small sample. Usually think aloud method is followed by a follow-up interview to ensure a complete set of data and gain clarification from participants(Fonteyn, Kuipers, & Grobe, 1993).

2.1.4.2 Research Methods

The previous section presented a variety of data collection techniques applied in the relevance literature. These techniques could be used in different research methods. This section elaborates on the various research methods that have pursued in relevance literature.

User Studies

User studies are research studies that aid in understanding user's needs, preferences, behaviour, opinion and evaluation (Sridhar, 1995). In relevance literature, user studies have been used widely to gain more understanding of the user's judgment

behaviour. More specifically, why and how relevance judgment decisions are made. In these studies, users are invited to participate in search sessions and their information relevance judgment behaviour was observed and analysed. Different data collection techniques were applied in these user studies.

Barry's (1994) study followed their relevance judgment user study with an open-ended interview. As the purpose of the study is to explore relevance criteria applied by the users, the un-structured open-ended interview was ideal to this aim. In this study, the participants were provided with a set of documents which are expected to meet their information needs, and they are asked to identify the parts of the documents that guide their relevance judgment decision. The role of the interview is to extract detailed explanation of their relevance criteria selection. Pre-defined questions did not exist. Wen et al. (2006) had also interviewed the participants before the actual search begins in order to predict the criteria users might apply in judging the relevancy of the retrieved documents. The main goal of the study is to examine the effect of topic familiarity on the number and type of resources on one hand, and on the relevance criteria choices on the other hand. After this initial interview, the participants were given two search tasks one is considered familiar and the other is non-familiar to them. A list of 12 criteria was provided to the participants and they were invited to discuss their relevance criteria selection after the completion of the searching sessions.

Similarly, Choi and Rasmussen (2002) have also applied interviews in their study of image relevance criteria. The interviews were structured using pre and post-test questionnaires. The aim of the initial interview is to investigate the users' perceptions of the importance of a set of relevance criteria without any search results. Then the authors conducted searches based on the search topics and keywords provided by the participants during the initial interview. The Library of Congress Memory photo archive was used and the search results provided to the participants so they can make their relevance judgments. Finally, follow up interviews were held using post-test questionnaires to rate the importance of relevance criteria again this time with

search results. Interviews have also been used in Hung et al. (2005) work, participants in this study invited to interviews after completing search session in order to clarify their relevance criteria selection.

The literature of relevance studies has shown some works which applied multiple data collection techniques in the same study. Both quantitative and qualitative methods were combined in the same study.

Sedghi et al. (2012) applied both semi-structured interviews and think-aloud protocol to collect data from health care professionals while searching for medical images. The interviews were not separated from the actual searching session, the participants were searching for images during the interview in order to clarify their relevance judgment behaviours. Grounded theory was used to analyse the collected data.

In her exploratory study, Hirsh (1999) have also applied interviews and think-aloud protocol to discover relevance criteria applied by children when searching for texts and images for a school project. More specifically, the author was interested in the information seeking behaviour of the children and the evolution of relevance criteria selections during the search process. In order to examine the change of relevance criteria choices at different search stages, the author conducted two interviews at different search stages. The first interview was in the first week of the research process while the second was in the third week.

The above-mentioned studies have applied two methods in collecting data however, another group of studies applied more data collection methods in their works. Xie and Benoit (2013) used both quantitative and qualitative data collection techniques in their study which aims to compare the list of relevance criteria of search result list and whole documents. The authors asked the participants to fill in pre and post questionnaires and to think aloud their thoughts while they are searching the web. Furthermore, log data have been analysed.

In another study, Maglaughlin and Sonnenwald (2002) asked the participants to fill in a questionnaire and to participate in an unstructured interview to gather background information about the participants. Then, after they engaged in a search session and judge the retrieved documents they were interviewed to gain a deep understanding of the reasons for participants' choices. Balatsoukas and Ruthven (2012) went one step further by adapting the eye-tracking device to gather the eye movements and fixation points of the participants. At the beginning of the study, the participants fill a background questionnaire then they were invited to a search session. During the session, an eye-tracking device collect data regarding the participants' fixations. Furthermore, talk-aloud technique was used to gather more understanding of the participants' judgment behaviours. Finally, semi-structured interviews were conducted. In these interviews, the videotaped of the participants' search sessions were presented and the participants were asked to clarify their judgment decisions.

Tombros et al. (2005) have also applied multiple quantitative and qualitative methods in their work which aims to investigate the relevance criteria employed by the web users. They specifically interested in the features that make a web document relevant, the effect of the task type on the criteria choices and the dynamic aspect of relevance criteria selections. Pre and post-questionnaires were used in addition to think-aloud protocol. Besides the before mentioned methods, the search sessions were recorded.

Similarly, Taylor (2013) has also applied three types of surveys: pre-test, in-test and post-test to gather data. The pre-test survey gathers background information about the participants while the post-test survey measures the user satisfaction level. The in-test survey asked that participants to rate the relevance of the retrieved document on a scale range from one to ten. The search session was followed by an open-ended interview to gather in-depth details about the reasons for judging documents as relevant or non-relevant.

Some of the studies let the participants conduct the search themselves, on the other hand, old studies were relying on mediator (usually the researcher) to conduct the search and provide the participants with results. This is might be because of the lack of user-friendly IR systems at that time.

Experiments

Part of the works in relevance literature follow an experimental approach where dependent and independent variables are identified, and hypotheses regarding the relation between both sets of variables are presented. Then, the experiment goal is to test the validity of these predefined hypotheses.

Examples of such works include the study of Xu and Chen (2006). They designed an experiment to test the validity of their five hypotheses of relevance criteria. The participants were given search tasks and asked to evaluate the relevance of two documents from the result list based on a semi-structured survey. More specifically, the participants were asked to rate to what extent each criterion contributed to their relevance judgment decision based on an eight-point scale questionnaire. Similar to Xu and Chen, Xu (2007) have applied the same method to test their hypotheses regarding the criteria that contribute more to relevance judgment in hedonic and epistemic information search.

Recall Studies

Another category of studies relied on participants recall their searching techniques and perception of the degree of relevance of the retrieved results without conducting a real searching session. One of the shortcomings of this method is that people forget things very easily and it is highly expected to miss important details by relying on participants' memory.

One of the early works that follow this direction is Schamber (1991). She interviewed users of weather information. The participants were asked to recall a decision making situations where the decision was based on weather information and all of the interviews were transcribed. Inductive content analysis was used to extract and classify relevance criteria mentioned by the participants.

The study of video relevance criteria of Yang (2005) has also based their findings on semi-structured interviews. The real search session was optional in this study. Similar to Yang, Savolainen (2010) conducted semi-structured interviews in order to investigate the criteria that people based their preferences of information sources in everyday situations such as “seeking for a house”.

The above-mentioned studies were gathering the data by interviewing participants. On the other hand, Ju and Gluck (2011) surveyed people perception of relevance criteria concept using questionnaires. They targeted the users of public libraries and asked 244 users to participate in a short questionnaire which asks to list three words come to your mind when you hear "relevance criteria’ phrase. The goal was to identify relevance criteria as perceived by participants and to distinguish core from peripheral criteria.

2.1.5 Relevance Summary

User relevance criteria have been studied intensively in text retrieval literature, however, the number of studies that focus on them in image retrieval contexts are obviously narrower. When it came to video relevance criteria, the number of studies that address them are quite smaller.

Thus, there is a lack of knowledge regarding how users apply relevance criteria in video relevance judgments. Few studies have focused on this issue, however; these studies did not consider all the different aspects of relevance judgments such as the various factors that might affect user’s relevance judgments. For example, the work of Yang (2005) has explored video relevance criteria in work task context. They

attempted to examine the influence of the different work tasks on relevance criteria choices. However, the study was limited to three different tasks and the tasks were work oriented. The author called for more investigation of the user video relevance criteria in order to generalize the findings. Furthermore, the study has been done more than ten years ago, users; relevance judgment behaviour might be changed since that time due to several reasons, one of them might be the appearance of popular video retrieval systems such as YouTube. Indeed, more work has to be done to gain more understanding of video relevance criteria outside this scope and this research is an attempt to fill in the gap of knowledge regarding video relevance criteria. As this research aims to study user video relevance criteria in a leisure context, a detailed review of leisure literature is presented in Section 2.2.

2.2 Leisure

The main goal of this research is to investigate the criteria users apply in making relevance judgment decisions when searching videos in a leisure context. In Section 2.1, background information of relevance concept and relevance criteria were presented, however, in order to gain comprehensive insight about this research it is essential to provide background information about leisure.

The remainder of this chapter is organized as follows. Section 2.2.1 introduces leisure concept and investigate its three different forms. Examples of leisure activites under each category of leisure are provided. Section 2.2.2 explains the relation between leisure realms and information behaviour studies. In particular, it describes how leisure been approached from information behaviour viewpoint. Consumption of image and video in leisure time is discussed in Section 2.2.3. Finally, Section 2.2.4 introduces leisure to relevance criteria studies.

2.2.1 Definition and Main Concepts

Human involved in various activities in their everyday life. Besides work and personal self-care activities, there are remaining free time through the day where leisure take

place. Leisure is defined by Stebbins as “uncoerced activity engaged in free time, which people want to do and in either a satisfying or fulfilling way (or both) use their abilities and sources to success at this” (Stebbins 2007, p.4). Besides being uncoerced, leisure is considered to be a positive activity, and thus boredom in free time is excluded from leisure as it is not a positive activity. By this definition of leisure which concentrates on the absence of coercion, Stebbins has contrasted previous literature that considers leisure as a "freely chosen activity" because he believes that there are many constraints that might affect the activity choices. For instance, "aptitude, ability and knowledge of available activities". People may want to do some activity but could not due to these constraints (Stebbins, 2009).

The above-mentioned definition have classified leisure as an "activity". The term activity is defined as "a type of pursuit, wherein participants in it mentally or physically (often both) think or do something, motivated by the hope of achieving a desired end" (Stebbins 2009, p. 620). Playing football, eating a meal, painting a roof and watching a movie are some examples of general activities. Activities can be classified as work, leisure or no work obligation. A more specific concept of activity is called core activity which is defined as "a distinctive set of interrelated actions or steps that must be followed to achieve the outcome or product that the participant seeks" (Stebbins 2009, p. 620). For example, the core activities in a gourmet cooking are finding a recipe and obtain the ingredients. The core activities in more casual volunteer services (such as, handing out leaflets or directing traffic) are the actions required to accomplish the service. In general core activities require individuals to take several interrelated actions to accomplish the activity.

The importance of leisure in people life lies in its contribution to their feelings of happiness and as a source of joy. “In surveys, many individuals value leisure above all else in life, on par with the well-being of family and home” (Hartel 2003, p.229). Stebbins pioneered in defining the serious leisure perspective (SLP) which is the theoretic framework that bridges and synthesises three forms of leisure. In other

words, "it is a way of looking at leisure activities and how people experience them" (Hartel, 2013).

The name of the perspective indicates that it is only about serious leisure, however, surprisingly the perspective includes casual and project-based leisure besides serious. Stebbins justifies the reason of the perspective label, according to him the name was simply related to the history of research in leisure which began early in 1974 on serious leisure. Casual leisure and project-based leisure emerged in later years. Furthermore, even that the name of the perspective was taken from the first type, this should not indicate that it is the most important form of leisure(Stebbins 2007). Jenna Hartel has illustrated the SLP by a diagram as shown in Figure 2.1.

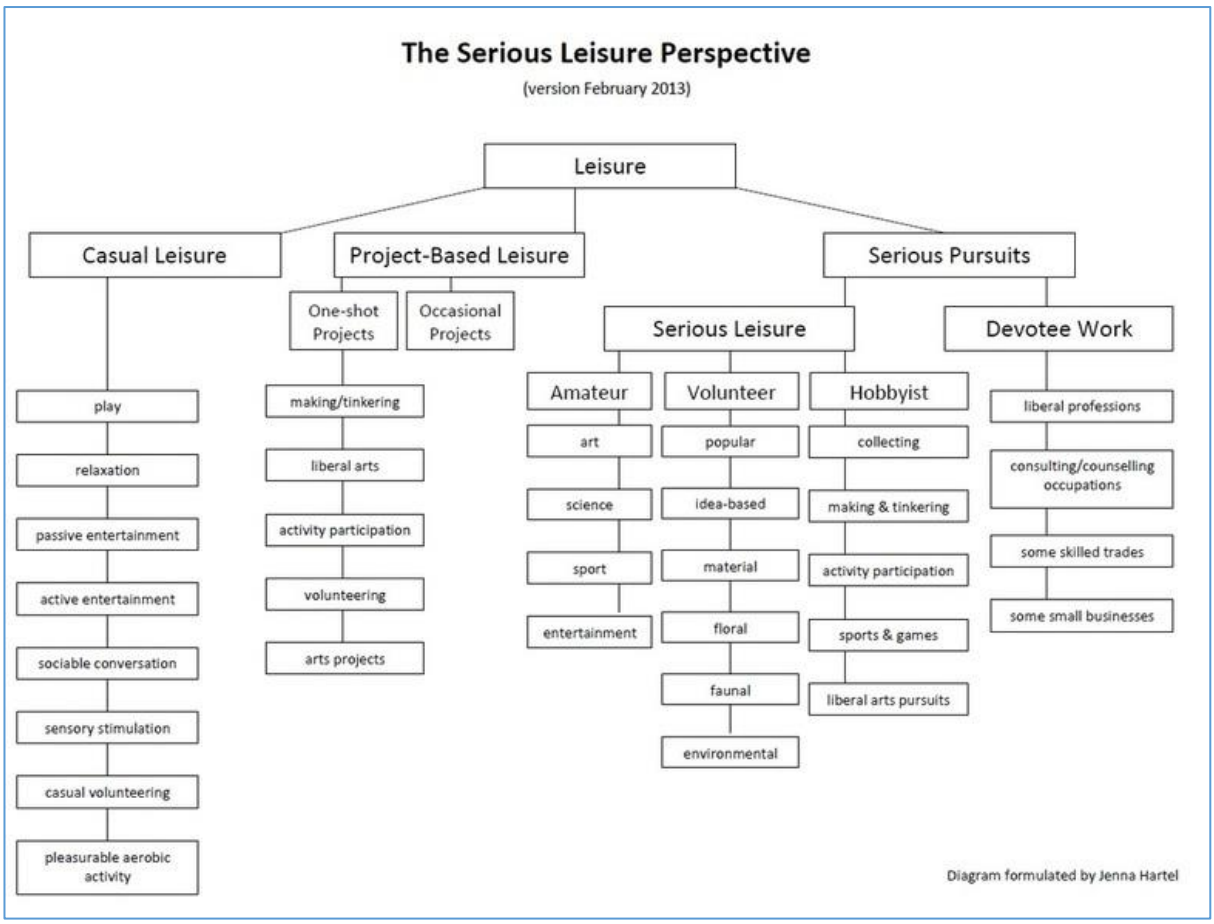

Figure 2.1 Serious Leisure Perspective (SLP) (Hartel, 2013)

According to Stebbins, the perspective attempted to bridge the gap that is known to separate the fields of leisure studies and library and information studies (LIS). It

simply introduced leisure activities and their different types to the researchers who are focusing on the information retrieval and behaviour of people. In general, the perspective serves as a facility for "systematically exploring people's use and dissemination of information during free time" (Stebbins, 2009).

The definitions of the three types of leisure as defined by Stebbins are listed below:

- Serious leisure: "the systematic pursuit of an amateur, hobbyist, or volunteer core activity that people find so substantial, interesting, and fulfilling that, in the typical case, they launch themselves on a (leisure) career centered on acquiring and expressing a combination of its special skills, knowledge, and experience."

• Casual leisure: "an immediately, intrinsically rewarding, relatively short-lived pleasurable core activity, requiring little or no special training to enjoy it."

• Project-based leisure: "a short-term, moderately complicated, either one-shot or occasional, though infrequent, creative undertaking carried out in free time. It requires considerable planning, effort, and sometimes skill or knowledge, but for all that is neither serious leisure nor intended by the participant to develop into such" (Stebbins 2009, p. 622). More details of each type of leisure are presented in the following sections.

2.2.1.1 Serious leisure

As the definition of serious leisure indicates, serious leisure activates demand "proactive acquisition of knowledge and skill" (Hartel, 2005). They can be classified into three main categories: amateurism, volunteering, and hobbies. Under each category, a set of subcategories are identified.

As illustrated in the SLP diagram, amateurs are found in art, science, sport, and entertainment. It is defined as "a member of professional-amateur-public system of functionally interdependent relationships, an institutional location that is both cause and effect of their serious, committed orientation toward the activity in question" (Stebbins 1982, p. 258).

The "public" term mentioned in the above definition refers to "a group of people with a common interest which are served by professionals or amateurs or both, and which make active demands on them"(Stebbins 2007, p. 586).

In other words, amateurs are individuals who decide to engage in a certain activity because of its strong attraction and for its durable benefits. Moreover, they do not find it enough to remain just a player or dabbler, they show more seriousness and commitment regard the activity and make more effort. However, amateurs are not a synonym of professionals. Two main differences are: First, professionals usually depend on the chosen activity to gain more than half of their income while amateurs not. Second, professionals usually spend more time on the activity than amateurs (Stebbins, 1977).

Furthermore, amateurs are distinguished from a hobbyist in that amateurs have professional counterparts and there is a strong relation between them and their professional counterpart, the activities that they engaged in considered work roles by other people. Examples of amateurs are genealogists and lace makers (Stebbins, 1982).

Turning now to hobbies, which is the most common category of the serious leisure activities. A hobby is defined as “the systematic and enduring pursuit of a reasonably evolved and specialized free-time activity”(Stebbins, 2017, p.72). According to Stebbins, a hobbyist in contrast to amateurs lacks a professional equivalent. In some cases, a commercial counterpart exists but without a relationship and interaction between the hobbyist and their professional counterparts (Stebbins, 1980).

Hobbies are classified into five subcategories: collectors, makers and tinkers, activity participation (in non-competitive, rule-based, pursuits such as fishing and skateboarding), players of sports and games (in competitive, rule-based activities with no professional counterparts like Chess and Monopoly), and the enthusiasts of

the liberal arts hobbies (e.g. acquiring knowledge of certain topic such as history or cultures) (Stebbins 2007, p.8, Hartel 2003).

Collectors are individuals who consider collecting objects as a hobby and not for any other profits benefits, the objects might be stamps, coins or bottles. The second category of hobbyists is makers and tinkers. Examples include pursuits such as gardening, cooking and knitting. The third kind of hobbyist is the activity participations where participants gain skills and knowledge while pursuing the activity. Examples include diving, fishing and bird watching. Players of sports and games is another type of hobbyists, examples of such activities include country running, softball and rock climbing (Hartel, 2005).

The last category of hobbyists is the liberal art enthusiasts who are people passionate with “the systematic acquisition of knowledge for its own sake” examples might be people who like a subject and spending time of self-learning and reading about it. Some of them go farther in their knowledge acquisition by engaging in cultural tourism and watching documentary videos and television programs (Stebbins, 2009).

The third class of the serious leisure is volunteering. A volunteer is defined as “one who performs, even for a short period of time, volunteer work in either an informal or a formal setting” (Smith, Stebbins, & Michael, 2010, pp. 239-240). The volunteer offers “uncoerced help” by providing other people (other than his own family) a service or benefit and usually without any financial rewarding. Stebbins emphasized on "the felt absence of moral coercion to do the volunteer activity" as a core principle in the leisure conception of volunteering. On the other hand, the non-profit sector research considers the absence of payment as a key element of the volunteering conception. According to Stebbins, there are two main motivations for volunteering activities: altruism and self-interest. Many individuals like to offer services to other people and have this altruism feeling, however, they only choose activities that contribute to their self-interest (Stebbins 2007).

According to Stebbins, all the forms of serious leisure (amateurism, hobbyists and volunteering) are common in six essential characteristics. First, serious leisure requires an occasional need to *persevere*, such as a basketball player who keep training to improve his skills. Second, serious leisure activities include acquiring knowledge and skills. The third quality is finding a career in pursuing serious leisure activity. Fourth, serious leisure has many durable benefits "which are personal and sociable rewards". Fifth quality of serious leisure is the unique ethos or culture that is related to all its forms. Finally, participants in serious leisure are identified strongly with their chosen pursuit (Stebbins 2007).

2.2.1.2 Casual Leisure

Casual leisure has been coined by Stebbins since 1982. Initially, it was used to sharpening our understanding of serious leisure by differentiation between what is considered serious or casual leisure. However, later on, researchers realized the importance of casual leisure as its own. It is the type of leisure which pursued by people much more than serious leisure and that is because individuals need to rest and recharge themselves from time to time by pursuing some casual leisure activities. Moreover, casual leisure considered the "main source of serendipity", it gives the participants the opportunity to experience and discover new things by themselves. Serious leisure activities also involve exploration and discovery, however, they are more *systematic* and *pre-arranged* (Stebbins, 1997).

Furthermore, casual leisure has a strong relationship with economics. Leisure industry has focused on many casual leisure pursuits and has an enormous economic income. Although serious leisure pursuits, for instance, golf, tennis and sport have gained the attention of leisure industrials, the size of industries that focused on casual leisure field, such as TV, is much greater (Stebbins 1997).

Types of casual leisure

Casual leisure activities can be classified into one of the following eight types: play, relaxation, passive entertainment, active entertainment, sociable conversation, sensory stimulation, casual volunteering and pleasurable aerobic activity.

Play is usually children-related activity, however, many adults' casual leisure activities are considered playing activities. Some of these casual leisure activities are taken seriously by other participants. Examples football, piano playing, playing with pets, video games and card playing.

Turning to relaxation which is defined as "release from mental or physical tension; especially by recreation or rest" (Stebbins, 1997). Sitting, strolling, napping, bubble bath and yoga exercises are all examples of this type of casual leisure.

Passive entertainment is a pleasurable activity that requires no effort from participants other than preparing to deliver it (for instance; turn the TV on, open a book). Various activities belong to this casual leisure category, for example, watching TV, reading book or magazine, listen to music and surfing the internet.

In contrast to passive entertainment, active entertainment involve action from participants. Museum and art galleries visiting, going to the cinema, shopping, riddles and puzzles are all examples of active entertainment. The main feature which is common to all these examples of active entertainment is the needless of a certain level of knowledge, skill or experience. However, if the participants need any of them, the activity is better classified as a serious leisure activity (amateur or hobby).

Moving to the sociable conversation, it is one of the forms of casual leisure where participants experience enjoyable playfulness activity. It can occur any time during participants waking time and it can be planned (such as gathering) or coincidence (chat with a co-worker). There are many examples of this category of casual leisure

such as: spending time with family, phone chatting with Mom, meeting friends, chatting with a co-worker during break time and communicate with others via social media (Facebook, Twitter).

Sensory stimulation is another type of casual leisure experience by everyone. Human engaged in various activities that demand sensory (smell, touch, see). Examples of these activities are: eating, drinking, listening to bird singing and observing birds and animals.

Casual volunteering is another type of casual leisure. As it is mentioned before, volunteering could be related to all the three forms of leisure: serious, casual and project-based. Casual volunteering is distinguished from serious volunteering by the needless of a significant level of experience or skills. Examples of it are: handing out leaflets, collecting donations.

The last category of casual leisure is pleasurable aerobic activity. It is defined as "physical activities requiring effort sufficient to cause marked increase in respiration and heart rate"(Stebbins, 2004). When it is pursuit regularly, aerobic activities could contribute in solving problems of obesity and diseases of the circulatory system. It helps in converting the boring exercises to more attractive and enjoyable ones and as a result, it contribute positively to the health and well-being of both children and adults. Examples of this type of casual leisure may include the following: bike rides, walking, treasure hunts game and Wii video games (which is produced by Nintendo and can detect the 3-D movements of the player).

It is significant to note that all the above-mentioned types of casual leisure are usually interrelated, participant of a certain leisure activity might experience two or three types of casual leisure at the same activity. For example, an individuals who likes to ride his bike (aerobic activity) may at the same time put on his headpiece and listen to music (passive-entertainment activity).

Stebbins stated that the main feature which is common among all the types of casual leisure is hedonic. More specifically, people engaged in casual leisure activities because of the enjoyment or pleasure feelings these activities produce to them (Stebbins, 1997). Stebbins has recognized five benefits of involving in casual leisure activities: (1) serendipity, (2) edutainment, (3) regeneration or re-creation, (4) maintenance of interpersonal relationships and (5) well-being (Stebbins, 2007).

2.2.1.3 Project-based leisure

Project-based leisure is the form of leisure coined by Stebbins. Prior to it, leisure has been thought to be even serious or casual. It is the form of leisure which is less common than serious and casual leisure. Stebbins identified the nature of participants who most likely intending to engage in project-based leisure activities as "people with heavy workloads; homemakers, mothers and fathers with extensive domestic responsibilities; and unemployed Individuals who, though looking for work, still have time at the moment for (mostly one-shot) projects"(Stebbins, 2005).

Similar to serious leisure, project-based leisure needs skills, knowledge and more importantly, a need to persevere. However, the main difference that distinguishes it from serious leisure is "failing to generate a sense of career". Moreover, project-based leisure activities are less complicated than serious ones. Project-based leisure activities might be more attractive to participants than serious leisure activities because of the lack of need for long commitment and participants could choose the suitable time to engage with it.

2.2.2 Leisure and Information Behaviour Studies

Although leisure occupied an essential part of everyday individuals' lives, it had not been intensively investigated by information behaviour researchers. Many reasons have contributed to this ignorance of leisure contexts in Information Retrieval and behaviour studies, one might be the typical perspective towards leisure as a trivial and marginal activity. However, participants in some leisure activities (especially

serious and project-based activities) do not consider these activities as trivial and marginal ones (Stebbins, 2009). Another reason for this neglect is the belief that leisure activities do not require information.

In contrast to this belief, Hartel (2006) stated that all the forms of leisure described before require to some extent information seeking and behaviour. More specifically, serious leisure activites, such as climbing mountains or lace machining, mainly depend on seeking some kinds of information. Depending on the nature of the activity, some casual leisure activities (excluding trivial casual activities e.g. napping and strolling) might also need the participants to seek certain information.

The main focus of the previous studies in Information behaviour were task-based scenarios usually in academic and professional contexts. Hartel justified the reason for this intention as the nature of these contexts. Academic and professional contexts tend to be more structured and accessible comparing to other fuzzy everyday contexts, in addition to the intensive amount of information required in them (Hartel 2003).

Later on, researchers in the information behaviour and seeking field start to draw more attention towards contexts other than professional and academic, where it is called everyday life information seeking (ELIS). However, these studies have investigated situations where accessing and retrieving information is essential to the user such as searching for information about breast cancer. Situations, where users are searching for their own fun (leisure context), have poorly investigated (Hartel, 2003).

Leisure context is worth investigation in order to examine whether the qualities of information change when moving from task-based scenarios in work context to leisure context. More specifically, leisure participants' information needs and the factors that have an impact on those needs have to be identified. Moreover, users of leisure information systems might be different of those who used information

systems in work contexts and they might lack the experience of using information systems outside the leisure situation (Elsweiler et al., 2011).

Bearing in mind that leisure has three main forms, research which has the intention of introducing the leisure realm to information behaviour studies follow the leisure perspectives. Hartel as an example attempted to investigate the information needs and seeking aspects of serious leisure activities such as cooking gourmet. On the other hand, other researchers (e.g. Elsweiler, Wilson) tend to explore the information behaviour in casual leisure situations. As the aim of this thesis is to investigate relevance criteria for videos in leisure context and because viewing videos online considered as a casual leisure activity (passive entertainment type), I will report on casual leisure studies only.

Casual leisure Studies

Elsweiler et al. (2011) explored the information behaviour of two different casual leisure scenarios. The first focuses on the use of television and the information behaviour associated with this use and the second was the casual information behaviour of Twitter users. In the first study, the goal was to explore the information needs and motivations of television viewers. Participants were asked to write down diaries describing their needs and the factors that motivate these needs. Then, an inductive grounded theory was applied to the collected information in order to analyze them. The result revealed a variety of needs with a different level of complexities. There were simple and well-defined need such as “How old was Tina when that concert was filmed?” other needs were still simple but less well defined, for example, “a list of interesting films showing from 7 to 8”. An example of a more complex need is “I’m looking for up to date news, I need to know the channel and time of broadcast”. In other situations the need is fuzzy and poorly described, examples, “I am looking for short entertainment during dinner”, “I’m zapping around again without a goal”.

Turning now to the motivations as reported by the participants, it had been found that a lot of motivations written down in the diaries is less dependent on finding certain information. Participants reported a variety of motivations such as, killing the time, distract their attention and provide entertainment. The experience itself was the main concern of the participants. For instance, a participant who is looking for a TV program just to “kill the time” main focus is not to find information.

By analyzing the motivations that trigger participants' information needs, the authors identified several motivational factors: user motivated (personal interest, knowledge or lifestyle, habits); context motivated (mood or state, time-related, socially motivated) and planning.

The second study focuses on self-reporting information seeking behaviour collected from Twitter. A corpus of tweets containing search keywords such as ‘search’, ‘browse’ and ‘explore’ were collected over five months. Example of such tweets includes: “Searching the net for birthday gifts ideas-getting really dump ideas”. Inductive grounded theory analysis has been applied in order to investigate the collected data. The Analysis revealed that many searching sessions have been identified as "needless browsing" where participants have no certain information need but they just would like to kill the time. For instance, one of the collected tweets says "I’m not even doing anything useful, just browsing eBay aimlessly. Furthermore, as with the previous TV diaries study, many participants are exploring for the experience. For them, the experience itself is more important than information finding.

Based on the findings of the two studies, four conclusions about casual leisure scenarios have been provided as follow: 1) the motivations of casual leisure searches are usually to reach particular mood 2) information needs are vague or might be totally absent in casual leisure searches 3) meeting these needs is optional 4) the experience of the search is more valuable than the information found. Elsweiler et al.

(2011) illustrated these four main differences that distinguish casual leisure scenarios in a revised information behaviour model as shown in Figure 2.2.

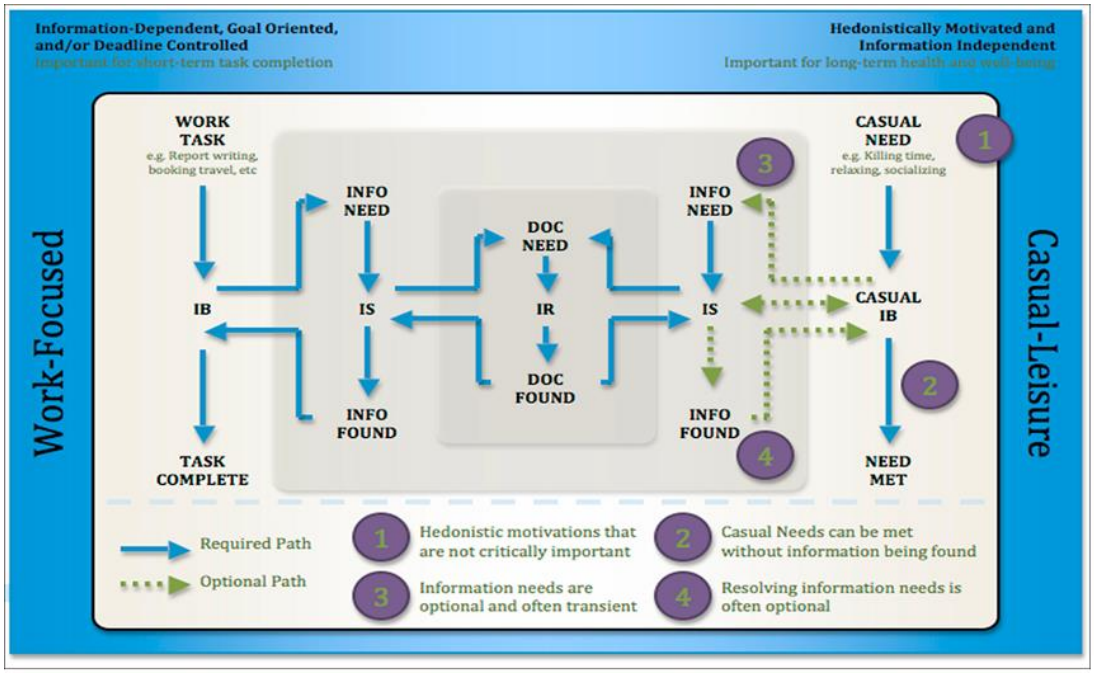

Figure 2.2 Casual-Leisure Behaviour Model (Elsweiler et al. ,2011)

In another attempt to explore casual leisure searching behaviour, (Millan-cifuentes, Myrhaug, & Macfarlane, 2014) recruited 28 participants to engage in casual leisure searching session motivated by curiosity rather than a specific information need. First of all, the participants were asked to fill in a pre-questionnaire, then they were provided by loosely defined simulated search task where the information need is fuzzy. In particular, the participants were provided by the following search scenario: “while you are waiting for your friends, explore what is happening in your city or other parts of the world”. The application used in this study were Twitter and Ambiecities. The latter is a social media application based on spatial-temporal data. There was no time limit for the searching session. Finally, the participants were asked to fill in a post- questionnaire to evaluate their searching experience.

The authors found that the participants using Ambiecities spent more time searching than the participants of Twitter. They justify the reason for this variation in search

session time by the motivation of the search. Participants who did their searches using Ambiecities are driven by curiosity and they keep engaging even when the retrieved tweets are not relevant. The design of the Ambiecities application facilitates an enjoyable searching experience for the participants. Another finding is: when participants are asked to identify the most important criteria that contribute positively in making their searching experience enjoyable, 71% of Twitter users choose topical relevance as the most important criteria which made the displayed information enjoyable. On the other hand, 21% of Ambiecities participants mentioned topical relevance. This finding shows how users of Ambiecities application, care more on the experience of using the application than the retrieved information itself. Ye and Wilson (2014) attempted to classify the searching session as leisure or work task, based on the behavioural data and the time of the day (e.g. early morning, evening). Twenty participants were asked to review their search history and classify the session to dimensions one of them was the importance of the sessions. Then, the study analysed the sessions which were classified as either high or low importance to predict the nature of them (leisure or work task). Based on the time of the day and the number of the pages viewed, many findings were presented such as viewing more pages during the day and the evening might indicate a casual leisure session.

Knäusl (2012) proposed a log- diary hybrid approach to uncover affective aspects of searching Wikipedia in leisure contexts. The proposed study aims to investigate users' motivations for searching, their information needs, satisfaction and emotional response to the searching experience. The ultimate goal is to learn what makes the users pleased during their searching experience.

A group of studies investigated users' information needs and behaviours in casual-leisure reading context. Wilson, Alhodaithi, and Hurst (2012) conducted a diary study to investigate user' information needs and motivations for casual-leisure reading. Twenty-four participants completed diaries in which they asked to make an entry for each reading activity conducted for leisure purposes. The findings revealed that the

motivations for casual-leisure reading were hedonistic or emotional based rather than informational based in around half of the reported motivations.

While Wilson investigated the information needs and motivations for leisure reading, Ross (2000) explored how readers select books for leisure reading purposes in physical libraries. 194 readers were interviewed by Ross and her postgraduate students to investigate the selection factors that influence the readers' choices of pleasure books. The reading experience wanted by the reader, elements of the book such as the book size, clues of the book such as title and publisher and accessibility of the book in terms of cost and time were found the main selecting factors.

In the same context of pleasure reading, Pöntinen and Vakkari (2013) integrated an eye-tracking system to elect eye movements for 30 participants when they select fiction for four search tasks in online public library catalogues. The goal of the study is to analyse the association between participants' fixations to metadata elements and selecting an interesting book. The results showed that content description and keywords gain the most attention from the participants but have no effect on selecting an interesting book. On the other hand, author and title elements gain less attention from the participants but significantly contributed in selecting an interesting book.

In another work, Mikkonen and Vakkari (2016) investigated fiction readers' interest criteria when selecting books in library catalogues for several search tasks. Eighty fiction readers were interviewed to express their selection criteria while they examining books using two library catalogues. The findings indicated five main dimensions of interest criteria: familiarity, bibliographical information, content, engagement and sociocultural criteria. Familiarity and bibliographical information are the most frequently mentioned interest criteria for selecting novels in both library catalogues.

Vakkari's studies applied interest as the notion for the criteria of selecting leisure fictions rather than relevance. This is because the authors interpret relevance as

topical relevance ignoring the users' emotions and assume a topical relation between users' needs and the retrieved information. Saracevic's model of relevance acknowledges broader aspects of relevance which goes beyond the topical dimension and includes situational, cognitive and affective aspects of relevance (Saracevic, 2007b). I believe that relevance is a more useful and comprehensive concept for this research context. In this research context, users' judgments decisions' regarding the retrieved videos for leisure purposes are not only based on interest, other criteria such as the quality aspects of the videos and topical relevance are important as well. Moshfeghi and Jose (2013) explored the characteristics of four different search tasks scenarios each with different intent: seeking information, re-finding certain information, entertainment by adjusting arousal and entertainment by adjusting mood. The goal of the study is to investigate the cognitive emotion and interaction aspects at different stages of the search process. Video retrieval system was used, 24 participants were asked to find as much as possible relevant videos to accomplish the tasks given to them. The findings showed that different search tasks with different intentions have different characteristics of cognition, emotion and interaction. More similar to our study context, Yeh (2016) explored the casual-leisure information behaviour of viewing videos online. Twenty-four undergraduate students were asked to complete diaries followed by interviews related to their online video viewing. Users' casual leisure information behaviour were investigated at three phases of the search: pre-viewing, viewing and post-viewing which results in a proposed framework of casual leisure video viewing processes and information behaviours. In the pre-viewing stage, the study reported on the motivations that triggers video search and classified these motivations to active seek for information (e.g. passing time, satisfying curiosity) and passive (e.g. information encountering or friend's recommendation). During the viewing stage, the findings showed that people seek information to satisfy cognitive and emotional needs. Finally, in the post-viewing stage, people behaviours include searching for further information and share founded information with friends.

The following section will discuss the role of images and videos in leisure searches. The discussion indicates why it is important to investigate relevance criteria for searching for videos for leisure purposes.

2.2.3 Image and Video: How are they related to Leisure?

Years ago, people used to look for information manually in text form in books and newspapers in order to satisfy their information needs. Information retrieval systems were mainly used in a professional context to solve work-oriented tasks and they were not available beyond this context. In their free time, individuals engaged in leisure activities such as watching TV or listening to the radio to satisfy their hedonic needs.

However, the invention of the Internet and the affordability of low-cost technology have changed the world. A PC device and internet access are all that an individual needs to be connected to the world. Search engines such as Google, Bing and Yahoo became available freely for millions of users. As a result, individuals could satisfy their information needs by using these affordable information retrieval systems. Moreover, the internet revolution has also influenced the entertainment and leisure sector. Surfing the Internet became one of the casual leisure activities people could do in their free time. Information retrieval systems are no longer exclusively used by professionals for work-related tasks, they have been available for the public to take advantage of them during leisure time as will be highlighted bellow.

Today is the era of image and video. As the internet speeds and bandwidth expanded, the visual content online has increased too. Online visual content has been used effectively in many fields and they succeed in delivering messages to recipients. As an example, the education sector has benefited from the availability and the possibility of retrieving and sharing visuals content online. Many research examined and encouraged the use of videos in classrooms to improve the students' learning outcomes (Berk, 2009; Duffy, 2008, Wu, Krajcik and Soloway, 2001). These studies

provide guidelines for using available videos and embedding them in teachers' presentations in the classrooms to attract the attention of the students and aid the ease of delivering information to them. Images and videos have also been used in marketing, brands use many social visual media in the advertisement process e.g. Instagram, Facebook (Carey, 2014).

Among the various ways images and videos have shaped our lives, the most interest of this research is the role visual content have in leisure and entertainment time. If "A picture is worth a thousand words then one minute of video is worth 1.8 million", so say Forrester's researchers (McQuivey, 2008). According to Cisco, video content will dominate the internet in the coming years. In particular, video traffic acquired 73% of all consumer internet traffic in 2016 and this percent is expected to be 82 by 2020 (Cisco Public, 2017). Adobe Digital Index has also reported in a study that focuses on videos on the web a significant rise in video consumption between the years 2011 and 2012. More specifically, more than 15 billion video streams measured in 2012 which means 50% growth in video consumption since the beginning of 2011 (Adobe Digital Index, 2013). The study has also reported an increase from 26% to 74% in Facebook viral reach (The number of unique people who saw a post from a story published by a friend) when the video is added.

Smartphones have also contributed to the increasing trend of retrieving images and videos in casual leisure time. The above-mentioned study of the Adobe Digital index revealed a rise in the mobile video consumption. Smartphones video consumption have tripled from 2011 to 2012 to gain 12% of the total video consumption. Cisco visual networking index predicted that by 2021 smartphones traffic will exceed PC traffic (Cisco Public, 2017).

This rapid growth of the visual content online indicates how individuals increasingly depend on images and videos to satisfy their information and entertainment needs. Variety of image and video sharing systems became widely popular today. Flicker, for example, an online photo management and sharing website that enables the users

to share their photos with friends or even the whole world. Flickr launched in 2004 and has 90 million monthly users around the world in 2018 (Smith, 2018). In 2010, another photo and video sharing application was launched with the name Instagram. In December 2014, the co-founder of Instagram Kevin Systrom announced that Instagram has 300 million users accessing the site per month. More recently, three Stanford University students: Evan Spiegel, Bobby Murphy, and Reggie Brown created a video sharing application called Snapchat. The application initially launched in April 2011 and has a special property that the 'snaps' are disappeared from the receiver device after certain time set by the sender. Snapchat had 191 million daily active users from around the globe (Statista, 2018).

Among the various video retrieval and sharing systems, YouTube gains special interest and today it forms one face of the modern media that cannot be neglect. Previous studies have shown that YouTube is the most popular video retrieval and sharing website (Lux et al., 2011).

YouTube is a video-sharing website that offers a diverse range of content and can be used in active or passive mode. In particular, using YouTube actively means creating and uploading content to YouTube. There are many YouTube channels today that consider different aspects of life. The contributors who upload contents to YouTube came from diverse sources and backgrounds, they might be "large media producer and rights owner (such as TV station, sports companies and major advertiser) or they might be small to medium enterprise looking for cheap distribution, artists and non-professional and amateurs media producer". Each of them uses YouTube with their own goals (Burgess & Green, 2009).

According to Karim (one of the founders of YouTube), four properties have contributed most to the great success of the site. The properties are: recommendation to the users by providing them with related video list, the email link which facilitates the sharing of videos, the comments which add social

communication to the site and the possibility of embedding YouTube videos in other sites (Burgess & Green, 2009).

The website was established in May 2005 by three former employee of PayPal Company: Chad Hurley, Steve Chen and Jawed Karim. Its goal was to facilitate sharing videos online by offering a user-friendly interface that makes video uploading, sharing and watching a straightforward process. Even users with limited technical knowledge could manage to use the website easily. YouTube has no restriction on the number of videos user can upload.

Shortly after its spread, YouTube was purchased by Google for US$1.65 billion in October 2006. Since then, YouTube became widely popular. In specific, according to Alexa official website (Alexa is a website owned by Amazon that measures the frequencies of visits on various websites), YouTube is ranked second globally after Google (Alexa, 2018). YouTube official website announced that the number of YouTube users exceed one billion around the world. Moreover, people watch one billion hours of videos daily and more than 300 hours of videos are uploaded to YouTube every minute. Half of YouTube content consumption occurs via mobile devices, a fact that compatible with the previously mentioned information about the role of smartphones in increasing trend of retrieving images and videos during casual leisure time. YouTube is hosted in 88 countries and available in 76 languages (YouTube, 2018).

YouTube has changed the rules of the games and media policies by offering the opportunity to all users in different communities to have their own channels and publish and spread their own ideas, products or talent with little restrictions e.g. copyright. This is an opportunity that was not afforded to these communities before and which lets video sharing became a way of everyday communication. YouTube offers an alternative for small enterprises or talented people to reach mass public. Amateurs could publish content online and reach viewers easily through YouTube.

Studies emerged to discover different aspects of users' behaviours on YouTube. The remainder of this section will explore some of these studies. Madden, Ruthven, and Mcmenemy (2014) conducted a content analysis of the users' comments on YouTube in order to examine and classify various types of comments. The study identified ten main labels with 58 sub-labels which indicated a wide range of uses of the YouTube comments feature. Halvey and Keane (2007) investigated users' search behaviours when searching videos on YouTube with the ultimate goal of comparing these behaviours to text searching. To achieve this goal, a sample of YouTube videos was built using a web crawler. The findings showed that users intend to interact with the services provided by the search engine and that tagging and textual descriptions contribute positively in making some videos more popular than others. As a result, researchers could benefit from text analysis techniques in video retrieval context.

Another group of studies has focussed on the intentions for viewing and sharing videos on YouTube. Hanson and Haridakis (2008) surveyed 251 college students in order to explore view and sharing behaviours for news content on YouTube. In specific, the study focussed on the factors influencing viewing and sharing of news videos and the motives that stimulate such activities. The findings revealed different motives for different types of news-related content. Yang, Hsu, & Tan (2010) examined users' motivations for sharing their own videos on YouTube. The authors went one step further by examining gender differences in sharing behaviours. They found that female users' intention is related to usefulness and social norms, while male users' intention is influenced by interpersonal norms.

2.2.4 Leisure and Relevance Criteria Studies

As mentioned before, leisure situations had experienced neglect from information behaviour studies. Relevance criteria studies were not an exception, its literature clearly shows how the previous works in relevance criteria have focused on work-related tasks and more specifically in academic or professional context. Few studies considered everyday contexts e.g. (Savolainen & Kari, 2006). In this study, the non-

work context was applied by asking the participants to conduct a search on self-chosen topic for example hobbies and to think aloud their thoughts.

From those studies which consider contexts other than work, few of them have investigated leisure. Xu (2007) attempted to explore relevance criteria selected by users in non-problem solving tasks, where tasks are self-generated and the motivation of the search is to satisfy user's "hedonic" (search for fun) or "epistemic" (desire for knowledge) needs. The author creates hypotheses of the importance of certain relevance criteria to the two different types of relevance: informative relevance (motivated by epistemic need) and affective relevance (motivated by hedonic need). The hypotheses were tested by asking 113 users to search the web to find documents that satisfy their needs and then to participate in a survey. The findings indicated that Novelty, Reliability, and Topicality are the most important criteria that contribute to informative relevance while Topicality and Understandability were the key criteria in affective relevance.

Furthermore, there were some studies that explored relevance criteria in different media such as video, however, even these studies had considered video relevance criteria in a professional context. As a result, exploring video relevance criteria in leisure context is worth investigation in order to discover whether the same relevance criteria applied while moving from work to leisure context.

Barry and Schamber work is considered a core study in relevance criteria literature. Thus, moving back to the list of relevance criteria identified in this study and to the criteria mentioned in each work separately, one could predict which of these criteria might be applicable to leisure users. Table 2.4 identified these criteria and explain how they could be applied in video retrieval in leisure context.

Criterion/source of the criterion	Definition	Application in video retrieval in leisure context
Clarity / Barry & Schamber	"the extent to which information is presented in a clear and well-organized manner"	For a user who searches videos for fun, clarity might remain an important criterion. More specifically, voice clarity and video resolution are all essential aspects that might affect the relevance decision.
Currency / Barry & Schamber	"the extent to which information is current, recent, timely and up-to-date"	Users might be only interested in recent movies or video documentaries.
Tangibility / Barry & Schamber	"the extent to which information relates to real, tangible issues; the extent to which definite, proven information is provided"	Might be applicable to video retrieval system in a certain situation. For example, users who are looking for leisure documentary might find tangibility an important criteria.
Quality of sources / Barry & Schamber	"The extent to which general standards of quality or specific qualities can be assumed based on the source providing the information; the source is reputable, trusted, expert"	Might be applicable in certain situations, when the user is looking for documentaries regarding events happened in the past. In this case, the reputation of the producer has an important role in relevance judgment.
Accessibility / Barry & Schamber	"The extent to which some effort is required to obtain information; some cost is required to obtain information"	Users might restrict their search to only free available videos.
Affectivness / Barry & Schamber	"The extent to which the user exhibits an affective or emotional response to information or sources of information; information or sources of information provide the user with pleasure, enjoyment or entertainment."	As the motivation of the users, while searching videos in leisure context are to satisfy their hedonic needs, feelings of pleasure and enjoyment are essential to consider the retrieved videos as relevant.
Novelty/ Barry	"The extent to which the information presented is novel to the user".	Individuals might restrict their search for only videos that they had not seen before.
Relationship with the author/ Barry	"The extent to which the user has a personal or professional relationship with the author of a document"	Users of video retrieval systems in leisure context might prefer certain movie actors or directors and probably will use these preferences as a criterion to judge the retrieved video.
Time constraint/ Barry study	"The extent to which time constraints or deadlines are a factor within the situation"	users might look for short videos to watch during their short free time (e.g. short breaks)
Ability to understand/ Barry study	"The user's judgment that he/she will be able to understand or follow the information presented"	Users will exclude videos recorded in languages user cannot understand.

Availability of interpretation or explanation/ Schamber study	"A summary, interpretation, or explanation is available"	the existence of subtitle for the videos recorded in foreign languages might be an essential criterion
Time Frame/ Schamber study	"Information covers a specific time frame"	Might be applicable for users who are looking for historical videos of a certain period of time.

Table 2.4 Predicting relevance criteria for leisure context based on the previous literature

Some of these criteria could be borrowed from the previous works of text relevance criteria and applied directly in leisure contexts, others might applicable in certain situations such as Tangibility, Time frame. Besides the previously mentioned criteria which already identified in the previous relevance criteria literature, investigating video relevance criteria in leisure context will definitely reveal more criteria. The priority in applying relevance criteria in leisure context might also be different.

2.3 Chapter Summary

To sum up, this literature review has shown that, although user relevance criteria have been studied intensively in text retrieval literature, the number of studies that focus on them in image retrieval contexts are obviously narrower. When it came to video relevance criteria, the number of studies that address them are quite smaller.

Thus, as mentioned before, there is a lack of knowledge regarding how users apply relevance criteria in video relevance judgments. Few studies have focused on this issue, however; these studies did not consider all the different aspects of relevance judgments such as the various factors that might affect user's relevance judgments. For instance, studying relevance criteria in another work tasks or to take one step further, which this work aims to do, and investigate video relevance criteria in non-work contexts, in specific, leisure context.

Leisure is considered a significant part of individual's everyday life. Leisure concept is tightly associated with pleasure as people engage in several leisure activities with the goal of achieving happiness, fun and joy. Despite the importance of leisure for

peoples' life, relevance criteria studies mainly focused on academic or work-related context.

The affordability of technology and the Internet facilitate the use of different information retrieval systems not just for work-related tasks but also to use these systems in leisure time. This phenomenon has attracted research community recently and researchers started to identify leisure concept and to investigate user's behaviours in leisure context. Thus, investigating how people judge objects retrieved for leisure search is crucial to meet the users' needs and better support their leisure searches. This research is an attempt to enrich the current literature by investigating users' video relevance criteria in leisure context.

Chapter 3

Methodology

The aim of this research is to investigate relevance criteria use when searching for videos in leisure contexts and the dynamic evolution of these criteria between different search stages (selection and viewing). The previous chapter reviewed relevant literature and demonstrated how relevance criteria for video retrieval in leisure contexts were under-studied. This research has been designed to fill this knowledge gap. Specifically, the study sought to answer the following research questions:

- RQ1: What are the relevance criteria users apply when judging videos in a leisure context?
    - Subsequently, which relevance criteria are the most important when judging videos in leisure context?
- RQ2: To what extent do these criteria match the criteria mentioned in the previous literature of text retrieval and/or work task context?
- RQ3: What is the difference in employing relevance criteria between the selecting and viewing stages of video/leisure contexts search?
    - Subsequently, are there significant differences in applying relevance criteria between the selecting and viewing stages of video/leisure contexts search?

- RQ4: Do different research methods provide different or similar findings? Does the diary method provide different findings from recorded search sessions with interviews?

The chapter is divided into several sections. It begins with the overall research design (Section 3.1) followed by the data collection methods applied in the research (Section 3.2). Research ethics and quality are also presented.

3.1 Research Design

Research approach selection mainly depends on the study's research questions and what a researcher is trying to find (Silverman, 2017). The overall design of this research is mainly based on qualitative methods. Qualitative research focuses on “the meanings people attach to things in their lives” (Taylor, Bogdan, & DeVault, 2016, p.18). Silverman stated that qualitative methods are more suitable if researcher aims to understand everyday behaviour (Silverman 2017, p.13). Ellis as well stated that qualitative methods are used to study people’s everyday lives and discover the needs and motivations that trigger their information seeking (Ellis, 1993). As the goal of this research is to investigate relevance criteria which are defined as the reasons users give to judge their videos as relevant, qualitative methods are suitable to answer such questions which deal with the "what" e.g. What are the relevance criteria users apply when judging videos in leisure context?, and "how" these criteria changes among different stages (Silverman, 2017).

The literature review chapter explored different methods used to study relevance criteria, among these methods, a qualitative approach was chosen where relevance criteria are extracted from the participants’ utterances rather than given to the participants as in some previous studies (Taylor, 2013; Tombros et al., 2005). This approach investigates relevance criteria for video retrieval in leisure contexts without limiting the participants by a set of pre-defined list of relevance criteria. In addition, making assumptions on what might be considered as relevance criteria for leisure

context based on previous academic and work-related contexts is inadequate. Differences in the contexts of the studies might affect the relevance criteria people apply. Previous research has begun to show evidence that casual-leisure search is different from typical web search in terms of the users' intentions and searching behaviours (Elsweiler, Wilson, & Harvey, 2012). Designing the research in a away where relevance criteria could be extracted from the collected data rather than only asking the participants to rank the importance of a given set of relevance criteria was more appropriate as it aids in inductive discovering of criteria used in this new context (leisure) but also made use of those already mentioned in the literature.

Although qualitative methods were used in this research, the research approach is not purely qualitative. A quantitative approach was also followed in data analysis as content analysis research method was applied. According to Neuendorf (2017), content analysis is thought to be a qualitative research method by many people but it is not. Both quantitative and qualitative techniques are encompassed in the process of content analysis. For example, counting mentions of relevance criteria and providing tabulations with these mentions. Statistical analysis of significance was also applied in the analysis to figure out significant differences in applying relevance criteria between various stages of the search. More about the approach followed in data analysis and how it mixed between qualitative and quantitative approaches will be provided in Chapter 4 Section 4.1.

The video retrieval system used in this research is YouTube. Other video retrieval systems were considered such as Dailymotion and Bing, but a decision was made to restrict the system used to YouTube. Among different video search engines, YouTube is the most popular site which is ranked second globally after Google (Alexa, 2017) widely used by over a billion users and with a huge variety of videos.

3.2 Data Collection Methods

With the aim of investigating video relevance criteria in the leisure context and examining the differences in applying relevance criteria at different stages of the search process, two studies were conducted. The first study investigates relevance criteria for videos in a leisure context and follows a naturalistic approach in which a diary was used as the data collection method. The goal of the first study is to answer research questions RQ1 and RQ2. In the second study, the dynamic use of relevance criteria is examined. A more controlled approach is applied in the second study in which recorded search sessions followed by semi-structured interviews were used. The second study aims to answer research questions RQ3 and RQ4. The following sections provide details about both of the data collection methods used including justification for using the method in addition to the advantages and limitations of each of the methods. Figure 3.1 illustrates the overall design of this research.

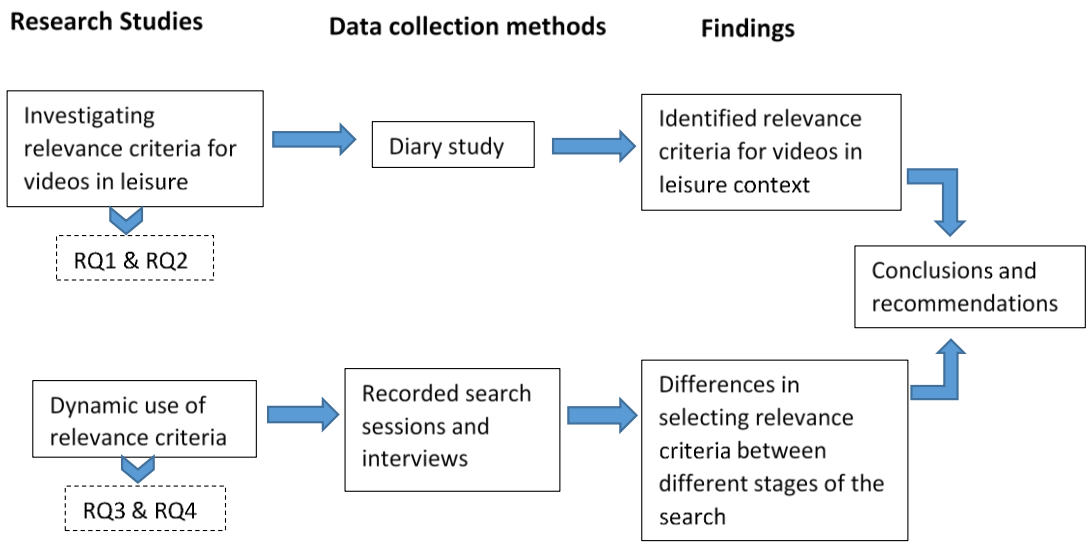

Figure 3.1 The research structure

3.2.1 Diary Method

The diary is defined as “a document created by an individual who has maintained a regular, personal and contemporaneous record”(Alaszewski, 2006, p. 1). Diaries have been widely applied by many disciplines such as health care, psychology and marketing (Götze, Prange, & Uhrovska, 2009; Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2012). Early research (Rieman, 1993) advocated the use of diaries to study user behaviour in HCI field and considers it as a method to balance the limitations of laboratory and field studies. In information science, diaries have been employed by researchers to investigate areas such as users’ information needs and seeking behaviours (Athukorala, Hoggan, Lehtio, Ruotsalo, & Jacucci, 2013; Elsweiler et al., 2011) or to study the motivations of information seeking using social media websites such as Twitter (Elsweiler & Harvey, 2015).

A diary is a natural self-reported instrument which is capable of collecting data in the users’ real environments without the influence of an observer (Elsweiler, Ruthven, & Jones, 2007). Users are required to repeatedly report about specific experience which allows the researcher to investigate the users’ everyday situations (Bolger, Davis, & Rafaeli, 2003). Therefore, it is a good research instrument choice for research that aims to understand the “why” aspects of user behaviours, in my case: why the user selects a specific video to watch. Unlike interviews, diaries overcome recall problems by minimizing the time between recording the event and the event itself (Bolger et al., 2003; Corti, 1993).

Although there are a number of advantages of diaries over other data collection techniques, diaries have some limitations. First, filling a diary can be a time-consuming task that needs an effort from the participants. Therefore, the diary period needs careful consideration. In addition, a successful diary study needs a good commitment from the participants who should be convinced to complete the diaries (Pickard, 2013). Another issue is that participants might miss valuable inputs because they thought it is not sufficiently interesting to record them (Sohn, Li, Griswold, &

Hollan, 2008). Thus, the level of detail required has to be clarified to participants at the beginning of the study (Palen & Salzman, 2002). Following a naturalistic approach using diary method reduced the level of control on the study as the researcher cannot directly observe the relevance judgment process. Thus, the analysis was mainly based on what the participants say in their diaries and their abilities in expressing their relevance criteria.

Other methods applied in relevance criteria literature were considered such as survey and talk aloud protocol. Survey is a well-known low-cost method to collect data from a large sample (Pickard, 2013). However, applying survey in this research will limit the participants' to a given list of relevance criteria and minimise the opportunity for new criteria to emerge. On the other hand, talk aloud used a lot in relevance criteria studies. Talk aloud aid in understanding participants' thoughts and procedures they follow to solve tasks. However, I believe talk aloud is not suitable to leisure contexts as it is an artificial process and my aim is to follow a natural and real approach as possible. More about talk aloud is presented in Section 3.2.2.2.

All in all, the diary was chosen as a research method because of its suitability to the research problem. Leisure is known to be an everyday life activity, thus, an instrument that is capable of collecting data in a more natural and realistic way is needed.

3.2.1.1 Types of diary

Diaries are classified into three main types:

- Interval-contingent design which requires participants to have an input to their diaries at pre-determined intervals.
- Signal-contingent design requires the participants to have an input to their diaries when they are prompted using signal devices at fixed or random intervals.

- Event-contingent design requires the participants to have an input each time the event under consideration occurs (Bolger et al., 2003).

The first two types of diaries are considered as time-based designs where researchers should carefully consider the most suitable interval for assessment. The participants' role is to report on their experience at a specific time (e.g. every three hours) or when notified. Leisure searches could occur anytime during the day and cannot be specific to certain time or interval, thus, time-based designs were excluded and event-contingent was the selected design for this study.

Bolger stated that applying event-based design, the participants need to be clear about what exactly counts as "triggering events" to avoid losing useful data. In this study, the participants are encouraged to write in their diaries for a duration of one week each time they search or browse videos on YouTube at their leisure times. They could record leisure sessions more than once a day with a maximum of ten sessions for the whole period of study. The number of search sessions and the number of videos to report on in each session were determined based on the diaries best practice which advised that the amount of information participants asked to provide should be carefully considered to avoid the burden of the participants. Furthermore, the suitability of these numbers of the sessions and videos was also checked through the pilot studies (Section 3.2.1.4) preceded the main diary study. These pilots showed that three videos in each session and maximum of ten sessions per participant are capable of collecting the required data. To ensure that participants were clear on what to report and the level of details required, they were provided by a sample diary taken from a pilot study.

3.2.1.2 Developing the diary design

As diary method success mainly depends on participants' commitments and satisfaction, careful design must be planned to achieve the study goal. Various design factors should be considered while designing the diary such as the time needed to

complete it, the duration of the study and the amount of information required. Rieman stated that diary study duration should not exceed 14 days, otherwise it might increase the burden of the participants (Rieman, 1993). In this study, I limit the duration to one week where participants choose any week that better suit them. The reason behind this decision is to follow the good practice of applying diaries and limit the risk of participants' boredom. In addition, the pilot study showed that one week is enough to capture useful data and it will cover different days of the week (weekends besides work days). Having a long duration for the study or asking for too much input will have the risk of participants' boredom and as result increase the probability of participants' dropout. Regarding the amount of information required, the diary entries mainly focused on the reasons participants give for judging the video as relevant or not. In addition, information regarding the context and motivations were also required. Thus, the goal of the study could be achieved with brief diaries.

Mainly there are three different ways of how inputs to diaries organized: structured, semi-structured and unstructured. Structured diaries are the ones which requires minimal typing more similar to the structured questionnaire with the difference of the frequency of asking for inputs. Questionnaires usually filled once while structured diaries can be considered as a questionnaire that has to be filled in repetitive manner. This way of designing the diary applied in some studies and called experience sampling methods however it is not suitable for this study as the goal of the study is to explore relevance criteria in video leisure contexts, so having pre-determined set of criteria and let the participant select what criteria he applied each time he searched for leisure is not an appropriate method. Thus, this choice was excluded. In contrast to structured diaries, unstructured diaries ask the participants to report on required activity freely without having a pre-determined structure. There is a danger that participants lose the goal of the study and fill out the diaries with useless details of their leisure session. Furthermore, it required heavy inputs which increase the burden and drop out of the study. So this option was also rejected. I believe that applying semi-structured design where participants are required to fill pre-determined fields with their relevance criteria will have the benefits of the two other

designs as it enables the participants to express their relevance criteria freely but at the same time ensures the study goal is met. In this case, the input is minimized and the opportunity to gain useful data is increased. At the same time, semi-structured design gives the participants the freedom to write about their reasons for selecting their videos in their own words without the restrictions of pre-determined list of criteria.

3.2.1.3 Diary instruments

Various methods could be used to record diaries. These methods are mainly classified into two categories: printed (paper and pencil) and electronic diaries. Paper and pencil was the earliest method used in diary studies. Using this method, participants usually provided with booklets to report their inputs. One of the advantages of printed diaries is the ease of use to the participants but it has many other limitations. The risk of forgetfulness, retrospection error and uncertain compliance increased using this method. Privacy is also an issue in printed diaries as the diaries could be viewed by other people. On the other hand, electronic diaries outperform printed diaries in the ease of data entry, management and accuracy. It also makes the analysis easier by eliminating the need to transcribe the data and also addressed the privacy issues by having a password protected forms (Bolger et al., 2003).

In this study, the participants were given the option to select from three different ways of recording the diary: printed, Word processor with a pre-defined template or Google document. These options were offered so the participants could select what suited them better. The printed diary option was piloted but no participants chose it in the main study. When used in the pilot, the diary was printed and handed to the participant who was informed to return it back upon completion. The second option is to send the participants pre-defined template of Word processor document and asked them to record their inputs using the file, at the end of the study the participants should send the file back by email. The last option is Google shared document. The option of using Google documents has the advantage of having live

access to the data while diary reporting is in progress. In case the participants misunderstand the goal of the study or provided information that lacks the required details, I contacted them by email to clarify the issue. In this way, better data could be achieved by discovering issues early on the data collection process.

3.2.1.4 Pilot studies

Pilot studies aid in minimizing mistakes that may occur in the main study. It is a way to test the instruments and bring out issues that might be costly to repair later (Teijlingen & Hundley, 2001). Two pilot studies preceded the actual data collection of the diary study. The aim of these pilots is to examine the ability of a diary, as a data collection tool, to collect the required video relevance criteria in leisure context and to improve the design of the diary.

To serve this aim, an initial semi-structured diary was designed (Appendix A.1). This early version begins with a short paragraph describing the goal of the study and giving the participants brief instructions of what should be recorded in the diary with some examples. Although the main goal is to investigate relevance criteria, other information regarding the context of the search was also required. This information include the time and location of the search sessions. This information could be used to discover when and where most of the leisure search took place.

Other information regarding the topic of the search or the motivation that triggered the search in case the participant was not looking for certain topic was also required. The aim of collecting this information is to have a better understanding of the context of the search session as mentioned above. The participants could record up to ten search sessions in their diaries for the duration of one week.

After revising the design of the first version, a modified version of the diary emerged as shown in Appendix A.2. In the modified version more detailed instructions were provided to the participants and less formal words were used. For example, 'videos

that you like' instead of 'relevant videos' and ‘reasons’ that make the participant like/dislike the video is used instead of ‘criteria’.

In addition, the first diary version was asking for too much input from the participants. For example, participants were asked to record their relevance judgment and the criteria applied to five videos. This has been changed to three videos in the modified version. The relevance judgment field has reconstructed as well, as an alternative of directly asking the participants to give their relevance judgment decision for each video, the participants were asked to record their reasons for liking three videos followed by the reasons that made them not willing to complete videos until the end.

The first pilot used the version in Appendix A.2. In this pilot, a form designed using Microsoft Word was used to collect the information. The participants received the diaries via emails and once completed they sent them back to me. Six YouTube users (four females and two males) volunteered to participate in the first pilot. Of the participants, two were PhD students, one engineer, one researcher, one computer courses trainer and one undergraduate student. The participants were young people in their twenties and they were from two countries: the United Kingdom and Saudi Arabia. All the participants filled their diaries in English in this pilot study. Preliminary analysis of the first pilot provides information regarding relevance criteria applied, leisure topics, average number of videos watched per session and the average number of sessions per participants. Some issues raised from the first pilot. For example, some participants were not comfortable with the diary in Word format. One of the participants reported that he did not get the meaning of ‘topic or motivation’ and recorded only the sessions when he was looking for a specific topic.

As suggested by Pickard (2013, p. 231), "the format of a diary can be predetermined, but very often you will need to amend this based on the reaction of participants to your design". Thus, a third version of the diary emerged and piloted. Based on the findings of the first pilot study, studies from the literature and participants' feedback,

slight modifications have been made to the diary design as shown in Appendix A.3. Besides providing the participants with English instructions of what should be recorded in the diary, translated paragraph in Arabic were provided to the participants whose first language is Arabic. A new field was added as well to ask about the device used in the search with the aim of investigating which device is more popular in leisure search.

As mentioned before, some participants found 'Topic/motivation of your search' field vague. Thus, topic and motivation were separated out into two fields. The form of the topic field has been changed to a question form and an example of topics have been added 'What are you looking for? E.g. muffin recipe'. A note indicated that the participant could skip this field if they do not have a certain topic to search for and move to the motivation field have been added. The same thing has done with the motivation field, the form has changed to question 'why did you start this search? E.g. kill the time, change my mood, I want a new dish for the dinner'.

Slight paraphrasing was made to the sentence which asks the participants to record their relevance criteria. Moreover, instead of asking the participants to record the reasons for stop watching some videos without completing them (which indicates that the video was not relevant) at the end of each search session, this field has been repeated to each video. Prior to start recording in the diary, participants were asked to fill a short demographic form which collects information about the age, gender, profession and frequency of YouTube searches.

Finally, more options of diary format were provided to the participants. Specifically, the participants have the choice to use Word document, Google document or printed diary. The second pilot used the version in Appendix A.3. Three YouTube users (one male and two females) volunteered to participate in this pilot. Two of them choose Google document and one printed diary. Of the participants, one was PhD students and the others were undergraduate students. They were studying different subjects: Information Science, Pharmacy and Business.

3.2.2 Screen Recording and Interviews

Many factors drive the selection of data collection methods, among these factors the nature of the data the researcher aims to collect and the type of questions that should ask to collect these data were the main factors (Pickard 2013, p.196). In this study, these were the factors that influenced the selection of the recorded sessions followed by interviews method. The goal is to investigate the reasons that underlie the participants' relevance judgments in leisure context at different stages of the search process. Thus, repeated questions about the reasons that made the participant judge the video as relevant or not were asked. This should be done with the aid of screen recording software to avoid the problem of participants' memory lapses.

Recorded search sessions have been used in previous studies in conjunction with other methods such as think aloud, questionnaire and interviews (Al-Harbi & Smucker, 2014; Y. Li & Belkin, 2010; Tombros et al., 2005; Wakeling et al., 2016). The interview is a well-known method which used widely in qualitative research (Bryman 2016, p.466). According to Pickard, "Interviews can be used for reconstruction of events, descriptions and feeling about current events and predictions of future developments" (Pickard 2013, p. 196). After conducting a leisure search session, the interview purpose was to follow the participant's relevance judgment process to discover the reasons that justify these relevance decisions. Some of these reasons (relevance criteria) include how the participant felt regarding the video being watched.

Investigating the dynamic aspects of relevance criteria in the first diary study was not applicable. Using a diary method, it is not practical to ask too much information from the participants (identify which stage they were in when applying certain criteria), as this might affect the response rate and increase the probability of increasing the participants' boredom (Rieman, 1993). As mentioned in Section 3.2.1, the diary should be designed in a way to meet the study goals without overwhelming the

participants. Furthermore, the data might not be accurate as participants might not be certain about these stages. The first study findings paved the way to further investigations regarding the dynamic use of relevance criteria as some responses revealed the stage of the relevance decision and how the relevance decision and relevance criteria selections might change as the participant move from selecting to viewing stage (more about this in Chapter 5 Section 5.5.4). However, another study was needed to fully investigate the dynamic change in relevance criteria.

While the diary used in the first study followed a more naturalistic approach, the method used in the dynamic use of relevance criteria study was more controlled from the following perspectives. The search sessions were recorded using screen recording software and there was a time constraint on the session length (20 minutes). Although the study has some controls, a balance between conducting natural and controlled study was considered in the study design. Participants were not given the same videos and asked to judge their relevance and there were no restrictions on the leisure topics they could search for. A natural approach where participants freely select what to watch is applied.

3.2.2.1 Interviews Types

Mainly there are three types of interviews: structured, unstructured and semi-structured interviews. Structured interviews are kind of questionnaire carried out by the researcher directly (Pickard 2013, p.199; Bryman 2016). The researcher is restricted to a predetermined list of questions with limited opportunity to interact by follow up questions that might expand the data. On the other hand, unstructured interviews is a "purposeful conversion" where questions emerge during the interview (Pickard 2013, p.200). Finally, the semi-structured interview combines properties of the two former mentioned types. The researcher prepares a set of questions called an "interview guide" with the flexibility of new questions to emerge as the interview progresses (Bryman 2016, p. 471).

The selection of the type of the interview depends on how the researcher approaches the research problem. According to Bryman, a semi-structured interview is applicable if the researcher has a clear focus and want to address specific issues. On the other hand, unstructured interview used when the researcher begins the investigation with a general understanding of the topic (Bryman 2016, p. 469).

In the dynamic use of relevance criteria study, semi-structured interview approach was applied as it is suitable to answer research question RQ3 which deals with the dynamic aspects of relevance criteria and how it might change at different stages of the search process. The reason for selecting semi-structured interview is that the investigation begins with specific issues to address which are the relevance criteria participants apply at different stages of the search process. In addition, the approach provides the required flexibility to respond and interact with the participant by asking follow up questions whenever interesting points are mentioned. At the same time, it minimizes the risk associated with unstructured interviews of getting out of the scope of the research because an interview guide is used to help in maintaining the focus on the study goals. Therefore, the main questions asked were the ones were pre-determined, however, dynamic questions asked during the viewing stage to elaborate more on the criteria used in this stage.

3.2.2.2 Alternative study designs

As the study investigates leisure context, the optimal study setting would be the place where the majority of leisure search occur within (which is 'home' as revealed by the diary study). Thus, I first considered collecting the data remotely by letting the participants record their leisure session at their own preferable time and place using screen recording software. Then I could interview them and watch back their sessions and discuss their relevance criteria choices at different stages. However, this option was excluded because the main goal of the study is to collect relevance criteria while or soon after the participants made their relevance decisions so the reasons for selecting or rejecting videos are still fresh in their minds. Having a time gap between

the search sessions and the interviews will increase the risks of memory lapses. In addition, following this strategy will add more effort to the participants by letting them install software, record and send their sessions, and asking them to participate in two activities (record the session and attending the interview).

Conducting interviews only without asking the participants to perform searches was excluded as well. Interviews could help in understanding what participants' information needs in leisure contexts are and based on what criteria they generally choose their videos. However, applying only interviews and asked the participants to recall recent leisure searches and what were their relevance criteria used in their relevance judgment decisions is impractical because of participants' memory lapses (Yang, 2005).

Finally, a decision was made to record the search session and make the interview directly after the session at the same time and place and to avoid sessions been recorded beforehand. I piloted slightly different versions of study procedures that I will explain in the pilot section.

The decision was made to give the participants the search scenario Appendix B.5 and prepare them to perform the search by a short pre-interview to put them in the context of the leisure search. In this brief interview, I had a chat with the participants about the motivations that trigger them to search YouTube and how they usually use YouTube for leisure purposes: what video content interest them (Appendix B.7). Having just a short chat as preparation to the search session rather than asking the participants to fill in a diary prior to attend the search session found more practical as it minimizes the effort on the participants and avoids unnecessary extra phases in the study.

3.2.2.3 Pilot studies

After excluding other options that need to collect some of the data before attending to the study location, I piloted several ways of collecting the data all in the same

session. Four volunteers participated in different versions of the pilot, two of them had participated in the diary study. I first considered collecting verbal responses while participants were conducting their search (talk aloud method) in which participants explain why they select a video to watch from result list and when they watched the video whether they found it relevant to what they want and why. I tried two slightly different procedures with two different participants, in one I was observing the participant while he was searching and talk loudly his relevance criteria. Both the screen and participant's talks were recorded. The participant found it inconvenient to be observed in this way. In the second attempt, I left the participant alone and asked her to explain her relevance decision as she progresses in the search. While the latter option minimizes the pressure on the participant, using talk aloud protocol in both procedures was found to be impractical in this study for two reasons. The task is to search for leisure or entertainment, and applying talk aloud protocol might spoil the participants' viewing experience. Also, in this study participants needed to be able to listen to the video while watching and it would have been impractical to ask the participant to talk while they are watching a video.

Therefore, applying talk-aloud was excluded and a decision was made to apply recorded search sessions followed by interviews. I piloted this method with two volunteers to decide on the suitable time limit to be set in the main study and the amount of useful data which could be gain from the search session. The first participant was given ten minutes to search while the second 20 minutes. After these pilots and discussing the design with the volunteer participants, a decision was made to set 20 minutes as the time constraint of the study. This time was found to be reasonable as it allowed the participant to view some videos with moderate length.

3.2.2.4 Recording and transcribing

All interviews were recorded using "voice memos" phone application then transferred to a password secured google drive. Transcription took place soon after each interview has been conducted.

Transcription is the process of converting recorded conversations into textual materials. It is a fundamental step of the preparation of data analysis as it provides the data in a textual format to be ready for analysis and "it represents what the researcher and transcriptions preserve" from the original audio recordings (MacLean, Meyer, & Estable, 2004; Miles & Huberman, 1994).

All interviews were transcribed in full. This includes all words uttered in addition to laughs, e.g. "It's not overwhelming [laughs]", uncompleted sentences, e.g. "I selected that one because it's…Normally I only had twenty minutes to see something about the volcano" and repeated words, e.g. "I don't know, I don't know basketball terms right now".

Transcriptions could be conducted by the researcher or other transcriptions. The advantage of having the transcripts typed by the researcher is that the researcher will be familiar with the data at this early stage of the analysis (Gray, 2014). On the other hand, transcriptions are known to be a time consuming and laborious process.

For this study, transcriptions were done by a native speaker transcriptionist. Using a transcription service was chosen for the purpose of time, as a non-native speaker took longer to transcribe than natives. Moreover, errors are more likely to arise when working with a second language, thus, to ensure the accuracy of transcriptions and avoid transcriptions' errors, I made use of paid professional transcription service (MacLean et al., 2004).

After receiving the transcripts, I checked all of them for errors or missing words or sentences. This was achieved by listening to the transcribed audio file and read the corresponding transcription to ensure its quality. I checked the transcripts to ensure that every utterance are mentioned and no utterances have been transcribed to wrong words. This process is known as spot-checking and is recommended when the transcription is done by another transcriptionist (not the researcher) and when interview participants are from different backgrounds (MacLean et al., 2004). Spot-

checking aims to ensure high-quality standards of the transcripts as they are the key materials used in the analysis. According to MacLean et al., (2004) the number of transcriptions to be checked depends on the study size and number of transcriptionists involved and it is not necessary to check all transcripts. However, I had checked all the transcriptions upon receiving them to achieve accurate transcriptions.

3.3 Research Ethics

It is mandatory for all research that deals with humans to granted ethical approval to ensure the participants' dignity and safety while conducting the research. Knowing that the research has granted ethical approval from the University will increase the participants' confidence in the research and the researcher (Silverman 2017, p. 58). The data collected in this research has granted two ethics approval from the Ethics Committee in the Computer and Information Science Department at the University of Strathclyde. Following the University's Code of Practice on Investigations on Human Beings, a consent form should be signed by the participants to indicate their agreement to participate in the study (Appendix A.4 and B.4). The goal and procedures of each study were explained to the participants using the information sheets shown in Appendix A.5 and B.3. As the research collects personal data on how participants decide on what to watch in their leisure time, participants were informed that the goal is not to judge what they select to watch and their data will be treated confidentially.

In specific, in the diary study, the participants were informed that their data will be kept confidentially and only the researcher will have access to their electronic diaries. When I report on the data, participants' names were replaced by assigned numbers such as P1, P2.

3.4 Validity and Reliability

Validity concerns with the credibility of the researcher's interpretations of the data (Silverman 2017, p. 384). Valid research should ensure that the research's conclusion matches the interpretations of the data and the researcher is not being selective on what to report on. Reporting the findings based on few well-selected instances of the data known as anecdotalism. According to Bryman, in qualitative studies researchers attempt to support their claims and conclusions by evidence in the form of quotes fragments extracted from interviews for example. There is a concern regarding the representativeness of these examples which might lead to the problem of anecdotalism (Bryman, 2016). One way to show the generality of an instance is to illustrate its occurrences in the data. Count of mentions of each relevance criteria was included in the findings of both studies with the aim of increasing the credibility of the claim and avoid the anecdotalism problem. This is known as 'using appropriate tabulations' method of which aims to more valid findings (Silverman, 2017).

In the diary study, I arranged short meetings with the participants to get their feedback and to clarify any ambiguities appeared in the diary to avoid miss interpreting the data. This procedure aims to increase the validity of the research findings.

Although triangulations have not been agreed on as a method of increasing the validity of the research (Silverman, 2017), it is recommended to apply triangulation to study phenomenan from various standpoints. In an attempt for data and method triangulation, relevance criteria were investigated at two different studies using different methods: diaries and interviews and by collecting two sets of data from different participants.

Reliability refers to the consistency of the data analysis either by the same observer at different occasions or by multiple observers at one time (Silverman 2017, p. 400). It is known also as dependability (Bryman, 2016) and it concerns with the degree of

transparency in reporting the research procedures so it could be repeated later. To achieve reliability in my research, detailed procedures on how the coding scheme was developed and the distinctions between different codes labels is illustrated in Chapter 4. Evaluating the quality of the coding process on the dimensions of *stability* and *reproducibility* were conducted. Stability examines whether the same coder assigns the same codes to instances throughout the coding process. It is also known as intra-coder reliability. To measure stability, several rounds of data analysis were conducted which allow me to check how consistence I was when assigning codes to the instances. Reproducibility examines whether different coders assign the same codes for the same instances. It is also known as intercoder reliability or investigator triangulation. To check the reproducibility dimension, another coder (PhD student) was asked to code a subset (10%) of the data collected for the diary study chosen randomly and Cohen Kappa inter-coder agreement test was conducted. More details about the inter-reliability test are presented in Chapter 4 (Lazar, Feng, & Hochheiser, 2017; Miles & Huberman, 1994).

3.5 Chapter Summary

This chapter has presented the research design overview. As mentioned in Chapter 2, various methods have been used to study relevance criteria.

In this research, two studies were conducted to investigate relevance criteria applied in video/leisure contexts. The first study followed a naturalistic approach in which diary was used as the data collection method while more controlled approach applied in the second study, recorded search session followed by a semi-structured interview. The next chapter will present the data analysis method and the process of developing the initial coding scheme used in analysing the data from both studies. The findings and detailed procedures of these two studies will be presented in Chapter 5 and 6. Chapter 5 presents the findings of the diary study and Chapter 6 presents the findings of the interviews.

Chapter 4

Data Analysis Method

This chapter aims to present the main data analysis method used in this research. Section 4.1 presents content analysis as the main analysis method applied. Then the chapter presents how the coding was accomplished and how the coding scheme was built based on previous schemes provided by the literature.

4.1 Qualitative Content Analysis

The main method applied for data analysis in both studies was qualitative content analysis. Other statistical analysis tests were applied for the recorded search sessions and interviews study that will be presented in Chapter 6. Qualitative content analysis is defined as "the making of inferences about data (usually text) by systematically and objectively identifying special characteristics (classes or categories) within them" (Gray, 2014, p.607). These categories might reflect on explicit or implicit communication (Hsieh & Shannon, 2005). This considered one of the unique properties of content analysis: the analysis could be applied on manifest or latent content. Manifest contents are visible and obvious components of the content. For example, themes and main ideas of the content. Mayring refers to explicit content as verbal material or description. Latent contents are the underlying meaning of the content or the conclusion (inference) to be drawn from the material (Cho & Lee, 2014; Graneheim & Lundman, 2004; Mayring, 2014).

Another characteristic of qualitative data analysis is that the analysis could follow deductive or inductive approach or a combination of both approaches. This is in term

of the generation of the initial categories to be applied in the analysis. Using deductive approaches the initial categories could be derived from previous literature or existing theory. On the other hand, applying inductive approach, the initial categories emerge from the data (Cho & Lee, 2014).

Another definition which emphasizes the classic quantitative approach of content analysis is the one provided by Silverman, “Content analysis is the process of establishing categories and then counting the number of instances under each of them” (Silverman 2015, p.116). Providing counts of mentions of each instance support the validity of the findings. The process pays attention to reliability as one of its requirement is to have a precise definition of each category in a way that other coders would be able to assign instances to the proper category and reach the same results using the same data (Silverman, 2015).

Mayring considered qualitative content analysis as a mixed methods approach including qualitative and quantitative phases. In the qualitative phase, codes or categories are assigned to text. These initial categories could be derived in deductive, inductive or both ways. In the quantitative phase frequencies of categories are counted (Mayring, 2014). The decision to follow a deductive or inductive approach of data analysis depends on the purpose of the study and what is previously known about the problem being investigated. An inductive approach is advised when little prior knowledge is available about the problem under investigation. On the other hand, a deductive approach is appropriate when relevant previous research exist or the aim of the study is to test a known theory or “retest existing data in a new context” (Cho & Lee, 2014; Mayring, 2014). In this research, deductive and inductive approaches of qualitative content analysis are applied. Miles and Huberman stated that coding through iterative cycles of deduction and induction power the analysis and aid the researcher to understand what is going on and why (Miles & Huberman, 1994).

As relevance criteria has a rich literature, it is worth to investigate what codes could be inherited based on the literature. Section 2.2.4 in Chapter 2 demonstrated an attempt to predict what codes will still be applicable in studying relevance criteria used when searching for videos in leisure context. This list of potential codes did not exactly match the final coding scheme resulted from investigating relevance criteria for videos in leisure context study and the dynamic use of relevance criteria study findings. However, it shows the potential codes based on the literature only and prior to any data collection and analysis. The goal was to predict what relevance criteria would still be applicable to leisure based on their definitions as stated in previous studies. Thus, a deductive approach enabled the use of these codes from previous literature as initial categories. Furthermore, one of the research aims is to examine whether the qualities of relevance criteria change when moving from academic or work-related contexts to leisure contexts. Applying a deductive approach of data analysis will aid in achieving this goal as it facilitates the comparisons between the two contexts.

Conversely, applying inductive content analysis facilitate the emergence of new codes from the data. These categories are more likely to be exclusive to the context being investigated. This will serve the main aim of the research: investigating relevance criteria applied for searching videos in leisure contexts. Thus, deductive and inductive approaches of qualitative content analysis were applied.

One of the important decisions when conducting content analysis is determining the unit of analysis (Graneheim & Lundman, 2004; Mayring, 2014). A decision should be made about the data that the analysis will focus on depending on the research questions. These selected data might be whole or part of interviews' transcripts, diaries or observations (Cho & Lee, 2014). For the investigating relevance criteria for videos in leisure context study (diary study), the unit of analysis was utterances in the "reasons" fields of the diaries. The unit of analysis for the dynamic use of relevance criteria study (recorded sessions with interviews) was interview text about the reasons for judging the video as relevant or not.

Mayring (2014) provides steps for both deductive and inductive approaches of qualitative content analysis which aim to clarify the procedures to apply when conducting qualitative content analysis. Following the deductive approach, the steps include:

1. Formulate clear research questions and describe the theoretical background
2. Defining categories from theory or previous studies
3. Defining the coding guideline including coding rules and examples
4. Preliminary coding the content
5. Revising the categories and coding rules after coding 10% to 50% of the data
6. Recode data based on revised categories
7. Analysis, counting frequencies of categories occurrences and contingencies interpretation

For the inductive approach, clear research questions still need to be determined as a first step. Some other steps are different or applied in different order as shown below:

1. Formulate clear research questions and describe the theoretical background
2. The determination of category and levels of abstraction
3. Coding the data and formulate categories
4. Revising the categories and coding rules
5. Recode data based on revised categories
6. Build main categories
7. Analysis, counting frequencies of categories occurrences and contingencies interpretation

The next section will discuss coding in details as it is a process involved in many qualitative analysis approaches such as content analysis and grounded theory.

4.2 Coding

Coding is an essential step in analysing qualitative data. It is known as the process of classifying qualitative data under different categories. This done by assigning labels or tags to chunk of words or sentences which represent a unit of meaning (Miles & Huberman, 1994). Coding is an essential phase of data analysis but it is not an equivalent to it. Data analysis is much more than assigning codes to data, it encompasses interpreting the findings and reflecting on the overall importance of them. Coding is a technique which helps the researcher to think about the meaning of the data. By coding, we mean the various ways of organizing qualitative data. Coding aims to sub-divide the data into meaningful categories which enable the researcher to understand the data and make interpretations and inferences. It mainly about linking raw data generated from qualitative research to concepts and enabling the researcher to differentiate and link between different fragments of data and reflect on these linked fragments. The data fragments which linked to each other should be about a specific topic or theme (Coffey & Atkinson, 1996; Miles & Huberman, 1994). The concepts that raw data are linked to identified beforehand from the theoretical framework or the research questions. Alternatively, the concepts might also emerge from the data themselves as mentioned above.

Coding could be considered as a way of simplifying and reducing the data to ease their retrieval (Miles & Huberman, 1994). This is in alignment with Baily's definition of coding

> "coding is the process of reading a copious amount of raw data, assigning descriptive labels-codes- to most lines of text, grouping the data based on relevant characteristics, and eliminating the chaff until the remaining portions are organized in such a way as to be useful for generating analytical insights" (Bailey, 2018, p. 161)

Beside data simplification and reduction, coding could be viewed as the process of data complication "it can be used to expand, transform and reconceptualise data opening up more diverse analytical possibilities" (Coffey & Atkinson, 1996, p.29).

4.2.1 Steps in the coding process

Different authors provide guidance on the steps of the coding process. Gray summarizes a few steps that should be applied in the coding process as part of data analysis that could be followed by any approach of analysis such as content analysis or grounded theory. First, the data should be in text format thus transcription should take place for audio recording resulted from interviews. The next step is familiarization in which the researcher starts to read the transcriptions to get a broad idea of what the data is telling without attempting to interpret the data. At this stage, general notes about the data could be taken as on points which look interesting or unique.

Following familiarization is the focused reading step where the researcher begins to read the transcripts in depth and underline keywords and making notes on the transcripts. This step is considered the beginning of the coding process. It is worth noting that coding should start as soon as some data are collected. The researcher should not wait until data collection has been completed to start coding. As coding at early stages of data collection familiarizes the researcher with the data. Bryman stresses on the same point that researchers should code the data as soon as possible to better understand the data and avoid being overloaded with the huge amount of text. Reviewing/amending the code is the next step where a second reading of the coded data should be done. Codes are reviewed and modified as needed. It advised at this stage to check duplication of the codes where two or more codes refer to the same meaning, and also to avoid duplication with codes previously mentioned in the literature. In this case the code mentioned in the literature would be used instead of generating redundant code. The final step is to find relations between the codes (Bryman, 2016; Gray, 2014).

Different researchers report on various types of coding. Mile and Huberman (1994) differentiate between three approaches for coding: priori, inductive and an approach which midway between them. Other codes types exist which mainly related to grounded theory approach such as open, axial and selective coding. The following section will present the coding process followed in this research resulted in building the coding scheme used.

4.3 The Coding Scheme

The coding scheme, or "membership categorisation device", which contains a collection of categories and a group of rules on how to apply data to these categories (Silverman 2017, p. 544). Building a coding scheme is an iterative process between establishing potential categories and testing these categories on data to see whether the categories can fit all the data.

As covered in Chapter 2, relevance criteria literature is rich and this resulted in various coding schemes emerged from relevance criteria studies. One problem of the immersive relevance criteria labels emerged from previous studies is the redundancy of many of them. Few attempts to consolidate relevance criteria resulted from the relevance criteria studies exist (Bales & Wang, 2006) however they covered few studies and are not context-specific. Another challenge in comparing relevance criteria studies is that various methodologies have been applied in the relevance criteria literature.

As mentioned in the previous section, to avoid unnecessary duplication of labels and to make use of the relevance criteria mentioned in the literature, a decision was made to build an initial coding scheme unified from various relevance criteria studies with the possibility of adding new codes as they emerge from the data. The process of building the coding scheme was done after collecting the data of the investigating relevance criteria study which will be presented in Chapter 5. The remaining of this section provides details on this process.

The process of building the coding scheme used in both studies of the research starts with an attempt to unify codes from different schemes. Labels of relevance criteria and their definitions from 17 studies were merged together in one table along with the name of the source study and examples of the code. The studies are: (Balatsoukas & Ruthven, 2012; Barry, 1994; Barry & Schamber, 1998; Choi & Rasmussen, 2002; Cool, Belkin, Frieder, & Kantor, 1993; Cunningham & Nichols, 2008; Maglaughlin & Sonnenwald, 2002; Reuter, 2007; Rieh, 2002; Savolainen & Kari, 2006; Schamber, 1991; Taylor et al., 2009; Vakkari & Hakala, 2000; Wang & Soergel, 1998; Xie & Benoit, 2013; Yang, 2005).

Relevance criteria that found to be not applicable to video leisure retrieval context were eliminated. For example, criteria such as "Consensus within the field" (Barry, 1994) or "Affiliation" (Crystal & Greenberg, 2006) which have an effect on relevance judgement in the academic context but did not apply in the leisure context.

Then I started to compare the codes and their definitions and group codes with similar meaning together. The following table mentions examples where definitions from different studies of two code labels (Recency and Novelty) are presented. The definitions were similar to each other's, so the definitions I applied in this research for Recency and Novelty were based on these previous studies. This is because the definitions inherited from the literature serve this research context and seen as suitable for judging videos in leisure contexts.

Code	**Source**	**Definition**
Currency/ recency	Barry & Schamber(1998)	The extent to which information is current, recent, timely, up to date
	Balatsoukas & Ruthven (2012)	How current, recent, or up to date a piece of the resource is
	Wang & Soergel (1998)	The comparative newness of a document with regard to the user's topic
	Rieh(2002)	Whether a document is up to date
	This research	The extent to which the video is recent and this is important to the user

Novelty	Barry(1994)	The extent to which the information presented is novel to the user
	Savolainen & Kari(2006)	The extent to which information provides new viewpoints or ideas
	Wang & Soergel (1998)	Whether or not the user has seen the document before or whether its content is new to the user regardless of when it was published
	This research	The extent to which the video or the information presented in the video is novel to the user

Table 4.1 Definitions of the same code by different studies

Besides the criteria which have the same labels but with slightly different definitions, some of the criteria mentioned in the previous coding schemes have different labels but with the same meaning such as Genre and Resource Type; Layout and Presentation; Audience, Intended use and Orientation/Level (Table 4.2). Thus, comparing and aggregating codes or codes' definitions from different studies aid in recognizing similarities and differences between the labels and reduce the number of previous codes to be considered in this research by eliminating duplicated codes.

Code	**Source**	**Definition**
Genre/type	Yang (2005)	The genre of a video: documentary, educational, feature films, etc.
	Reuter (2007)	Mention of the genre of the book
	Maglaughlin & Sonnenwald (2002)	The form or type of artifact
	This research: Genre	The extent to which the genre of the video (e.g. animi, historical, comedy) is a factor in the relevance judgment
Layout presentation	Xie & Benoit (2013)	The display of documents or information
	Yang (2005)	How the information was presented in the video
	This research	The extent to which presentation, delivery and clarity of the information are factors in participant's relevance judgment
Audience Orientation level Intended use	Yang (2005)	Whether the video was targeted at certain **audiences** the participant wanted to
	Maglaughlin & Sonnenwald (2002)	focus on information indicating the intended **audience**
	Wang & Soergel (1998)	At which intellectual level the document is written and for which **audience** it is intended
	Xie & Benoit (2013)	refers to the targeted **audience** of the document

	This research	Audience is not used as an independent code in this research. Responses related to a targeted audience are coded as coverage (depth) code

Table 4.2 Codes' labels with the same meaning

In case of multiple labels are available for the same criterion, I chose one of them. For example, Genre is applied in the coding scheme used in this research with a definition similar to Yang's video relevance criteria study. Layout/presentation code was applied in this research with a definition that concentrates on the presentation, delivery and clarity of the information as seen by the user. This definition was developed based on the previous studies and the data collected from this research. I believe that defining Layout/presentation label this way fit the related instances in the data. The Audience, Orientation Level and Intended Use labels were not applied as separate labels in this research. However, targeted audience was included in the Coverage label as it represents the level of details provided based on the audience. More details about this will be covered in Chapter 5 Section 5.5.1.

Another example of codes which have overlapped meaning but with different labels are the ones mentioned in Table 4.3. Reputation, Reliability and Quality are all related to the quality of the source providing the information. So for this research, Reliability, Reputation and Quality are merged to form one criterion named Quality of Source. This definition includes cases where quality is assumed by the users based on the author or uploader of the video (e.g. YouTuber).

Code	**Source**	**Definition**
Reputation	Xie & Benoit (2013)	The extent to which the source of a document or information is well known or reputable
Reliability	Savolainen & Kari (2006)	The extent to which a source providing information is seen as trusted or reputable
Quality	Balatsoukas & Ruthven(2012)	The reliability and quality of the contents, or the reputation of the author and the resource.
Quality of source	This research	The extent to which participant's judgment of the quality of the information is influenced by the source providing the video and whether the source is well known or trusted.

Table 4.3 Quality of source criterion

In other cases, I needed to form a new label that represents some merged codes. As an example, several codes related to the depth, focus and variety of information were used in the literature (Table 4.4). Coverage is a new label emerged which is based on the codes in Table 4.4 which were mentioned in previous studies.

Criteria	**Source**	**Description**
Depth/Scope/Specificity Specificity	Barry & Schamber (1998) Savolainen & Kari (2006)	The extent to which information is in-depth or focused, is specific to the user's needs, has sufficient detail or depth, provides summary, interpretations, or explanation, provides a sufficient variety or volume The extent to which information is focused enough to match the needs of the user
Scope	Balatsoukas & Ruthven (2012)	For judgments about the depth, scope, completeness, or level of specificity of information.
Depth	Xie &Benoit (2013)	The extent to which detailed information is provided by the document
Scope	Xie &Benoit (2013)	The extent to which information is covered.
Specificity	Xie &Benoit (2013)	The extent to which information covered by the document is focused to match the user needs
Variety	Savolainen & Kari (2006)	The extent to which the source provides a sufficient variety of information
Coverage(Depth, Scope, Specificity, Variety)	This research	The extent to which information gained from the video is detailed and has sufficient depth, specific to the participant's needs, provides a summary, or provides a sufficient variety or volume of information

Table 4.4 Merged codes to form "coverage" code label scheme from previous literature

Then, using this initial coding scheme, each utterance was assigned a preliminary code from the derived coding scheme where possible and the initial coding scheme was revised iteratively during the coding of the data. This iterative development of the coding scheme aided in tightening the definitions of each label mentioned in the scheme based on confusing cases found in the data where more than one code could be applied. Rules of what is included or excluded under each code were developed. For example, in some responses, the participants assume an acceptable level of

quality based on the existence of known people in the video (e.g. YouTuber). At the first glance, ones could be confused whether People in the video or Quality of Source should be applied in this case. A decision was made to code such responses as Quality of Source. The reason behind counting such responses under Quality of Source is that people in these responses are treated as a source providing the information and the participant's judgement decision was not simply influenced by their appearance in the video, however, their appearance in the video indicates some level of quality.

Prior to assigning codes to data, utterances that touched on more than one relevance criterion were split. For example, the utterance "It was in the suggesting video column and had over 2 million views. I had not heard the song before" was split into three text fragments and assigned a relevance criterion to each: Recommended video ("It was in the suggesting video column"), Popularity ("had over 2 million views") and Novelty ("I had not heard the song before"). In this case, the selection of codes (relevance criteria) to be applied to each utterance was mutually exclusive.

As mentioned in the previous section, the coding followed both a deductive and inductive approaches. The analysis process was iterative between developing the coding scheme and assigning codes to participants' utterances. Data which could not be assigned a code based on the initial coding scheme were reconsidered. This process is known as deviant case analysis (Silverman, 2017) or analytic induction (Bryman, 2016) and it aids in increasing the validity of the coding scheme. New codes label were generated if necessary to fit these data. In specific, Recommended video, People in the video and Habit were new codes added to the scheme. A full discussion of these codes with examples are provided in Chapter 5.

In some cases, code definitions could be adjusted to include these data. An example of adjusting code definition could be illustrated by Topicality definition. Initially, Topicality refers to instances where information provided in the video matches the participant's search topic. As cases from the data emerged where participants mentioning the topic of the video as a criterion for selection without having a precise

search topic, Topicality definition was altered to be: "the extent to which information provided in the video matches the participant's search topic or interest".

After coding the data, I compared instances which have assigned the same category to ensure the category represents them all and they all contribute to the same theoretical concept. This process is known as a constant comparison. Bryman defines constant comparison as "a process of maintaining a close connection between data and conceptualization, so that the correspondence between concepts and categories with their indicator is not lost" (Bryman 2016, p. 573). Constant Comparison method helps in achieving more valid findings by investigating and comparing fragments of the data related to a single case (Silverman 2017, p. 390).

Finally, the criteria (codes) were clustered into categories based on similarity. The grouping of relevance criteria was not straightforward. Different categorizations of relevance criteria appeared in the literature. For example, Yang and Marchionini (2005) used three categories to classify all the criteria: textual criteria, audio/visual criteria and implicit criteria. Maglaughlin and Sonnenwald (2002) used six categories to classify the criteria resulted from their analysis. The categories are: abstract, author, content, full-text document, journal/publisher and participant. In this research, the criteria were grouped into eight categories as will be discussed in Chapter 5 Section 5.5.1.

4.3.1 Reliability Measure

The process of content analysis pays special attention to reliability issues (Silverman, 2015). This is because of the nature of content analysis which focuses on ambiguous qualitative data that could have multiple interpretations, which increase the probabilities of biases and inconsistencies when interpreting the data. The same researcher might code the same data differently at different points of time. The same words might have different meanings based on the context and different phrases could refer to the same meaning. Moreover, different coders might interpret the data

in different ways (Lazar et al., 2017). Therefore, reliability control is a key goal in content analysis process to ensure the quality of the coding.

Various inter-coder reliability measures might be applied to ensure the quality and consistency of the coding. All of these measures require another coder to code part of the data then consistencies between the two coders is checked. One of the well-known reliability measures is the percentage of agreement between coders which simply calculated as the number of cases where both coders agree on the same code divided by the total number of cases. A limitation of this measure is it does not consider the fact that coders could agree on the same code for some cases by chance only. Therefore, the percentage of agreement measure is excluded and did not apply in this research. Cohen's Kappa is another reliability measure that overcomes the chance agreement limitation. The Kappa value (κ coefficient) could range between -1 and 1, where values less or equal to 0 indicate that all cases of consistency between coders are by chance, 1 indicates perfect agreement (Hruschka et al., 2004).

After conducting and analysing the data from the first diary study, an objective coder (PhD student) who was not involved in design or data collection of the study was asked to code a subset of the data. Around 10% (79 instances out of 787) of the total number of instances was chosen randomly to be coded by the volunteer coder. The coder was provided with full descriptions of all the codes and discussion was made to ensure that the coder understanded every single code and the differences between the codes' definitions. Cohen's Kappa inter-reliability test was performed as it overcomes the limitation found in the percentage of agreement measure and is suitable when the codes are mutually exclusive and two independent coders are involved. The κ coefficient of inter-coder agreement of the two coders found to be 0.70 which is considered as a substantial level of agreement (Hruschka et al., 2004; Neuendorf, 2017). According to Lazar et al. (2017), Kappa values between 0.60 and 0.80 interpreted as satisfactory level of agreement and values above 0.80 are near-perfect agreement.

4.4 Chapter Summary

This chapter has presented the main data analysis method applied in the research. Qualitative content analysis was applied with the aim of interpreting the collected data and investigating the use of relevance criteria in leisure/video context. The process followed to develop the coding scheme for the research presented in details. The next chapter will report on the findings of the first study of this research which investigated relevance criteria for videos in leisure context.

Chapter 5

Investigating Relevance Criteria for Videos in Leisure Context

This chapter reports on the findings of the first study of this research. The purpose of this study is to understand how typical users of YouTube judge the relevance of videos in leisure contexts; what are the reasons users give when judging video material as relevant or not relevant? It also aims to investigate whether relevance criteria applied for videos in leisure contexts are similar or different from the criteria mentioned in the previous literature of text and /or work task contexts. This latter aim will be covered in Chapter 7. The main findings presented in this chapter are also found in Albassam and Ruthven (2018). The chapter begins with a brief overview of the study (Section 5.1) then presents the sample used in the study (Section 5.2) and the detailed procedure followed (Section 5.3) and the data analysis process (Section 5.4). Finally, the findings are presented (Section 5.5).

5.1 Overview

This study applied diaries as the data collection method and it aims to address research questions RQ1 and RQ2. Justifications of the method selection and pilot studies which preceded this main study were presented in Chapter 3 Section 3.2. The final diary design applied in this main study is presented in Appendix A.3. As shown in the diary, the participants were asked to report entries up to a maximum of ten sessions. Session is defined in this study as the self-determined period of time where

the participant is browsing or searching for a consecutive set of videos. Sessions could be of any length of time and for as many numbers of videos, however, the participants were only asked to report on a maximum of three videos in each session. For three videos participants choose to watch, they were asked to provide the reasons that make them select these videos. If they stopped watching any of these videos and did not complete it until the end, they also required to provide the reasons.

5.2 Participants and Recruitment

The sample used in this study was a convenience sample from students at different universities. The recruitment took place in the United Kingdom and Saudi Arabia. In total, 30 YouTube users participated in the main diary study. Twenty of the participants were female and ten were male. The participants were young adults between 17 and 35 with an average age of 23. Among the participants, 18 were undergraduate students, five PhD students, three master's students, two college students and two unemployed graduate students. They were studying different disciplines at various universities. One limitation of this sample type is the age of the participants (young) and the level of education (well educated). Relevance criteria might be affected due to these factors.

In general, participants acknowledge themselves as heavy YouTube users. In total, 20 participants reported that they search YouTube daily and the remaining search YouTube one to four times a week. As the design of the diary did not change a lot between the pilots and the main study, the data collected from eight participants (six females and two males) from the previous pilot studies were also included in the analysis making the total number of participants 38. The participants were recruited through e-mails and flyers distributed at different universities and college (Appendix A.7). Five-pound shopping vouchers were offered to the participants. The full demographic data of the participants of this study is shown in Table 5.1.

P#	Gender	Age	Profession	Specialization	Method	Country	Nationality	Frequency of using YouTube
P1	F	24	Master student	information and library studies	Google doc	UK	British	Several times a day
P2	M	20	Undergraduate student	Computer Science	Google doc	UK	Chinese	once a day
P7	F	23	Master student	Finance	Google doc	UK	Chinese	1-2 times a week
P9	F	21	Undergraduate student	Information technology	Word	KSA	Saudi	once a day
P10	F	21	Undergraduate student	Information technology	Google doc	KSA	Saudi	Several times a day
P11	F	21	Undergraduate student	Information technology	Google doc	KSA	Saudi	3 or more times a week
P12	F	21	Undergraduate student	Information technology	Google doc	KSA	Saudi	Several times a day
P14	F	22	Undergraduate student	Information technology	Google doc	KSA	Saudi	Several times a day
P16	F	20	Undergraduate student	Information technology	Google doc	KSA	Saudi	1-2 times a week
P17	F	20	Undergraduate student	Information technology	Word	KSA	Saudi	Several times a day
P18	F	21	Undergraduate student	Information technology	Word	KSA	Saudi	3 or more times a week
P19	F	21	Undergraduate student	Information technology	Word	KSA	Saudi	Several times a day
P22	F	20	Undergraduate student	Information technology	Google doc	KSA	Saudi	1-2 times a week
P23	F	20	Undergraduate student	Information technology	Word	KSA	Saudi	Several times a day
P24	F	27	PhD student	Pharmacy	Word	UK	Chinese	3 or more times a week
P25	M	29	Undergraduate student	Physiology	Google doc	UK	British	Several times a day
P26	F	28	PhD student	Electronic and electrical engineering	Google doc	UK	Chinese	once a day
P28	M	26	PhD student	Biophysics	Google doc	UK	British	once a day
P31	F	28	PhD student	design	Word	UK	Italian	3 or more times a week
P33	F	20	College student	3D animation	Google doc	UK	British	Several times a day

P34	F	17	College student	Television, Commercial projects	Google doc	UK	British	Several times a day
P3	M	19	Undergraduate student	Medicine	Google doc	KSA	Saudi	Several times a day
P5	M	20	Undergraduate student	Engineering	Google doc	KSA	Saudi	Several times a day
P6	M	20	Undergraduate student	Engineering	Google doc	KSA	Saudi	Several times a day
P35	M	23	Undergraduate Student	electronic and electrical	Word	UK	British	3 or more times a week
P41	M	35	Master student	Digital media	Word	UK	Indonesian	3 or more times a week
P40	F	23	Unemployed graduate	Mathematics	Google doc	UK	British	Several times a day
P42	M	21	Unemployed graduate	Physics	Word	UK	British	Several times a day
P44	M	33	PhD student	Operations Management	Word	UK	Chinese	Several times a day
p46	F	20	Undergraduate student	Prosthetics & Orthotics	Google doc	UK	Chinese	3 or more times a week

Table 5.1 Participants' demographic information

5.3 Procedure

Individuals who responded to the flyers and email invitations had to contact me by email to express their interest in participating in the study. Then I replied to them by email that explained the details of the study aims and procedures and provided them with a link to a demographic form and another link for the diary (Google document). The demographic form presented an information sheet, which introduces the researcher and explains the study goals, followed by a consent form. In addition, the demographic form collected some demographic information about participants such as age, gender, profession and the frequency of searching YouTube for leisure or entertainment (Appendix A.6).

Each participant had their own Google document to record entries in, and this was shared with me only. In addition to the Google document, participants had the option to fill the diary in one of the following formats: Word document or printed diary. Thus, a word document was sent to participants who do not prefer Google document. Eleven participants chose the Word version, 19 the Google document and none of them chose a printed diary. Google documents were used because of its ease of use and I had a live access to the participants' diaries, so I could contact them when something was going wrong. Using Google documents was useful to help increase the response rate, as one does not have to wait until the end of the week to get the data back and one can always get fragments of the data in case of uncompleted diaries. To make the process of filling the diary more convenient and avoid language barrier that might affect the participant's ability to express their relevance judgment criteria, participants from Saudi Arabia were offered the option of maintaining their diaries in either Arabic or English. Two participants wrote their diaries in Arabic and one mixed between Arabic and English. For those three participants who wrote their diaries in Arabic, I translated the diaries into English prior to the data analysis phase. The participants were asked to keep recording in their diaries for a duration of one week and to make sure that they recorded the information while they searching YouTube or soon after to avoid any memory lapses.

To ensure that participants understand what kind of information they could record, a sample diary with three examples was attached. The participants were informed that the relevance criteria they could record are not limited to the criteria provided in the sample diary and that these examples were provided to clarify the type of information they can record and the level of details expected. According to Bolger et al. (2003), keeping in touch with the participants in a non-intrusive way during the process help in retaining the participants of a diary study. Thus, I kept in touch with the participants via e-mail to answer their questions, and encourage them to keep going with the work. After completing the diaries, I contacted the participants to set a time to meet and have a post-diary discussion where I could clarify any ambiguities

appeared in the diaries and to get the feedback of the participants and provide them with the compensation.

5.4 Data Analysis

Preliminary content analysis of participants' diaries took place at the early stages of the data collection process by careful reading of the diaries. Early analysis is recommended as analysing huge amount of data at the end of data collection period might make the analysis process overwhelming and demotivating which affect the quality of the results (Miles & Huberman 1994, p.50). As mentioned in Chapter 4 Section 4.3, the coding process followed an a priori coding approach because of the previous rich literature on relevance criteria. Following this approach, coding schemes were unified on the basis of previous relevance criteria studies including Barry (1994), Barry & Schamber (1998), Yang (2005), Savolainen & Kari (2006), Balatsoukas & Ruthven (2012). Full details about the process of the coding and the coding scheme are presented in Chapter 4 Section 4.3. In many cases, I needed to examine the exact video on YouTube to understand the content or discover names of YouTube channels, people or events.

5.4.1 Participants Feedback

After diary completion, I met the participants individually. In the meeting, participants were thanked for participating in the study and provided the shopping vouchers as compensation for their participation in the study. The meeting was also useful for clarifying any ambiguities that emerged in the diary. Participants were also asked for their overall participating experience in the study. The majority of participants reported a positive experience, for instance, some participants mentioned that filling out the YouTube diary led them to discover how they search YouTube usually. P1 said "It makes me notice myself more and think why I am watching what I watch. I learned more about myself which was good". Another example is from P25 who said:

"Thank you for letting me participate, I really discovered new things about the way I watch YouTube, for example, I discovered that I'm more likely to watch YouTube late in the evening than during the day and I'm more attracted to watch short videos (3 mints) that express new ideas".

The participants were also asked about the design of the diary and whether the right questions have been asked. Generally, the participants were satisfied with the diary design. For example, participant 12 mentioned "The questions were good and what I like about them that they were open questions so I feel free to write what I want without being restricted and at the same time they were simple and gentle", participant 16 "Yes, the questions were very clear and simple and the quantity of them were also appropriate so I did not get bored while completing the diary". Other early feedback from participants regarding the diary design helped in reconsidering the amount and details of the instructions given, however, the diary itself were not affected. For example, one of the participants mentioned that he assumed that videos should be watched in English only. As some of the participants are overseas students and the videos that they are going to watch are more likely to be in their mother languages, this feedback was useful and brought to my attention the importance of mentioning this point in the instructions. Thus, a sentence which stated that you can watch your video in any language but write your inputs in English was added to the email which contains the descriptions of the study. Another useful feedback was about the number of watched videos per session. One of the participants said, "It was assumed that each session would probably contain more than 3 videos, which is not the case for me as I watch a single one every once in a while". So because I have restricted the number of videos to record to three per session, the participant thought it is obligatory to fill all three. Thus, it was essential to clarify that participants could watch videos as much as they do usually but record a maximum of three videos.

Finally, the participants were asked whether the search sessions provided in their diaries represent the typical way of how they use YouTube. In general, all the

participants agreed that the diaries represent how they usually search YouTube, however, some participants reported that the number of sessions reported might vary according to the affordable free time the participant has. The questions asked in this short interview are provided in Appendix A.11.

5.5 Findings

In total, the participants provided 234 search sessions with an average of six sessions per participant (SD 2.7 min 1, max 10). The total number of videos recorded in the diaries was 579 with an average of 15 videos per participant (SD 8.2 min 2, max 30). The results reported in this section include responses from the pilot studies.

Ten search sessions were excluded because they did not meet the requirement of the research (such as searches related to course assignments). There were also cases (48) where relevance criteria could not be extracted from the reasons participants mentioned for selecting their videos, e.g. “No reason”, “Autoplay” these responses were omitted from the analyses. After cleaning the data, a total of 787 instances of relevance criteria were extracted from the diaries.

The participants most often watched YouTube on mobiles (37.2%), laptops (30.3%), desktop (19%) and tablets (3%). The majority of leisure search sessions were performed at home (77%) the remaining were at work (9%), University (7%), transportation, e.g. bus and car (3%), restaurants and coffee shops (2%) other places such as a friend’s house, beauty salon and gym (1%). In terms of the time of the day, leisure search sessions usually occurred in the evening (45%) followed by afternoon (30%). The fewest number of sessions was in night time (24%). There were some missing data because of the cases where participants did not complete these fields.

5.5.1 Relevance Criteria

In total, 28 relevance criteria were identified through the analyses of the diaries’ content and they were grouped into eight categories as follow: criteria related to the

information content of the video; criteria related to the participants' previous experience and background; criteria related to the participant's beliefs and preferences or their situation; criteria related to the quality aspects of the video or the source providing the video; criteria related to audio/visual features of the video; criteria related to the accessibility of the video; criteria related to other information within the environment; and criteria related to other people's opinions or YouTube's recommendations. Table 5.2 provides a summary of the criteria and their categories. The number of mentions of each criterion will be presented in Table 5.3 page 122.

Category	**Criteria**
Criteria related to the information content of the video	Coverage
	Topicality
	Recency
	Genre
	Length
	People in the Video
Criteria related to participant's previous experience and background	Background/ experience
	Novelty
	Familiarity
Criteria related to the participant's beliefs and preferences or situation	Affectiveness
	Serendipity/Curiosity
	Habit
	Time Constraint
Criteria related to the quality aspects of the video or the source providing the video	Quality of Source
	Content Quality
	Technical Quality
Criteria related to audio/visual features of the video	Cinematography
	Visual Appeal
	Sound/Voice
Criteria related to the accessibility of the video	Cost
	Language
	Version
Criteria related to other information within the environment	Availability
	Verification
	Unusualness
Criteria related to other people's opinions or YouTube's recommendations	Rank Order
	Popularity
	Recommended Video

Table 5.2 Summary of the relevance criteria grouped in categories

A detailed discussion of each criterion is presented in turn below.

Criteria related to the information content of the video

This group of criteria focus mainly on the information content of the video and was the biggest single category of criteria with 40% of instances of relevance criteria in this category. Six relevance criteria included in this category, what follows are explanation and examples of each of them.

1. Coverage (depth, scope, specificity, variety)

This is defined as the extent to which information gained from the video is detailed and has sufficient depth, specific to the participant's needs, provides a summary or provides a sufficient variety or volume of information. Depth mentions include examples such as "he's giving too many details" and "upload in-depth reviews". Moreover, responses related to the targeted audience of the video are also coded under depth code, for example, "the title seemed appropriate for beginners" and "I stopped watching as, although somewhat interested, found it to be rather entry level (in terms of content and data provided)". The reason for counting responses related to the target audience as depth is that the level of details will vary based on the audience the video targeted. In some cases, mentions of depth refer to how easy the video was to follow and watch. For example, "it's easy on mind" and "easy TV series to watch".

Scope deals with the breadth rather than the depth of the video. Responses related to the scope of the information include "I stopped because I got enough information" and "The title made me think this video is more comprehensive than the first. In football there are more skills than just scoring goals. Especially that Best played in midfield not in attack (that is something I found out on Wikipedia when I did search after the second video)".

Specificity concerns with how the video was specific to the participant's need. Examples include "I did not saw it all because it's presented a lot of products and I

am interested in certain product, so I only saw the part that I am interested in" and "The interesting part was only the first 15 seconds or so, because it was the same 15 sec that my friend sent me".

Responses related to variety include "I chose this one in particular as it seemed to cover a fair range of topics", "it was a mash-up of a heavy rock band and a folky band so it sounded like it could be entertaining" and "list with variety of pop songs".

Some responses related to the thumbnail of the video such as "the picture was showing a before & after makeup, which I liked, so I got to see the result before watching the whole video" and "Before and after picture" counted in the Coverage category because the thumbnail gives a summary of the video.

2. Topicality

Is defined as the extent to which information provided in the video matches the participant's search topic or interest. This was the category with the largest number of mentions in the data. Typical responses demonstrating the matching between the video's information and the participant's search topic are "That's what I was searching for", "The title was perfect. This was exactly what I was looking for". When the video did not match the participant's topic the participant might skip the video "I stopped watching as I found the video wasn't exactly what I wanted", "I thought this was the video I was looking for, but it wasn't. So I stopped in less than 5 seconds" or continue watching the video even when it is not on the topic "so I watched it even though it is not from the Baku race which was initially my reason for this search". Other relevance criteria contributed to the participant's decision to continue watching the topically non-relevant video. In this example the ranking order of the video and its novelty were the reasons of the relevance decision, "This video turns up to appear on top of the search results and I had never seen this crash before".

Other responses mentioned a topic that the participant wants to know more about e.g. “Trying to find a way to successfully install Rei’s Minimap Mod for the latest version of Minecraft”, "I want to learn more about how the effects of the series are created". Responses that describe the content or the aboutness of the video were also coded as Topicality, e.g. “Do pranks to his guests and also for the viewers, criticizing the reality of the society”, "addition to that, he makes funny comments about bad videos on YouTube". In some cases, the participants were predicting the video was going to match his/ her interest such as “the title is interesting”," sounded cool so I thought I would watch to see the effects".

3. Recency

Is defined as the extent to which the video is recent and this is important to the participant. Responses such as “uploaded recently”, “Looks like quite new” and “the newest video” are coded as Recency unless there is an indication that it is the Novelty (see below), rather than Recency, of the video is the basis of the relevance decision. For example, “the newest episode I have not watched it yet” is coded as Novelty rather than Recency because the main reason for selecting the video is that the participant has not seen it before.

4. Genre

Is defined as the extent to which the genre of the video (e.g. anime, historical, comedy) is a factor in the relevance judgement. An example of responses coded as Genre include “I don't have a specific reason, just I like watching anime and prefer anime videos among other videos on YouTube”, “I like rap music” and “I always like watching bloopers”. Genre does not necessarily match the YouTube classification of the video which seems to be more general than genre as it appears in this study data.

5. Length

is defined as the extent to which video length (duration) is a factor in the participant's judgement such as "Its full duration is 1 hour; that's a really long time", "I stopped watching as it was too long". Another criterion, strongly related to Length is Time constraint (see below).

6. People in the video

This defined as the extent to which the participant's judgement is influenced by people appearing in the video (TV host, singer, actor, band, YouTuber or guest, etc.). For example, "plus Leonardo dicaprio was on the video's picture", "The guest is a funny person so I guessed that this episode must be good to watch". In some responses, the participants assume an acceptable level of quality based on the existence of known people in the video (e.g. YouTuber). This type of responses was coded as Quality of Source (see below).

For example, "I was aware this YouTuber was one for providing practical instructions and tips". The reason behind counting such responses under Quality of Source is that people in these responses are treated as a source providing the information and the participant's judgement decision was not simply influenced by their appearance in the video, however, their appearance in the video indicates some level of quality.

Criteria related to the participants' previous experience and background

This group of criteria are related to how the participant approaches the video with respect to her previous experience and knowledge. More specifically, what the participant knows about the general topic of the video, whether she has seen the video before or how familiar she is with the source providing the video.

1. Background/experience or personal memories

Is defined as the degree of knowledge with which the participant approaches the video, as indicated by mentions of background or experience or personal memories.

Examples of background or experience include “I chose this video knowing very little of the subject, in fact, I knew nothing at all about whatever the ‘dark net’ was”, “After watching his goals I was interested in his personality. I wondered why such a successful player was an alcoholic. So I saw this video, which seemed as an interview from the picture, and opened it”. Examples of personal memories include “Memories of my own hamsters made me want to watch”. Background/experience means that the participant is familiar with the topic of the video or has personal memories around it. This is different from the criterion Familiarity (below) where the participant is familiar with the exact or similar video or the source providing the video.

2. Novelty

is defined as the extent to which the video or the information presented in the video is novel to the participant, which means it is new based on previous interests, examples include “Never heard of this song before so I want to listen to” and “I think there is nothing new, I know all what she said”.

Mentions which indicate the participant’s desire to keep up to date with a topic are treated as Novelty mentions. Examples include “To keep up to date with the latest gaming news”, “I like to keep up to date with movie news”. The reason for counting these mentions as Novelty is that the participants while they keep up to date with a topic, they are aiming to get novel information or videos based on previous interest. This aim matches Novelty definition.

Keep up to date also includes mentions of watching a new video from preferred show. For example “Unseen episode of an Internet Show I’m following”, “I realised that I hadn’t seen the previous week’s video so went to find that one as well “and “To check new events happened to this family”.

3. Familiarity

Is defined as the extent to which the participant is familiar with the exact video or similar videos or is familiar with the source providing the videos. The familiarity with the exact video means that the participant would like to re-watch a previously watched video. For example, “I have watched it before so I knew I would at least enjoy the song and listen all the way through”, “I have watched this video clip many times”. Familiarity with similar videos examples include “Fuelled by my surprising enjoyment from this previous video, I decided to watch this one as it was of a similar nature”, “I have watched other videos in this series before”. Examples of responses where the participant is familiar with the source providing the video include “Uploaded by a channel I subscribe to, I chose this video because I always enjoy their videos”, “I enjoy the Outsidexbox videos that describe and analyses games from all different eras”. When the familiarity with the source providing the video leads to an assumption of quality, the utterance is coded as Quality of Source.

Criteria related to the participant’s beliefs and preferences or situation

This group of criteria are not related to the participant’s background experience; rather they are related to her preferences, emotions or situation.

1. Affectiveness

Is defined as “the extent to which the participant exhibits an affective or emotional response to video; the video provides the participant with pleasure, enjoyment or entertainment or alternatively disappointment or other negative experiences” Savolainen and Kari (2006). Responses that included mentions of emotions such as boring, cute, humour, laugh, funny, not interested, happy, exciting are assigned this code. Examples from the diaries include “Funny, and has a great sense of humour”, “I found the song boring, didn’t match my taste in music” and “This video, uploaded by a channel I subscribe to was chosen as a means of escapism”. This was the criterion with the second largest number of mentions in my data.

2. Serendipity/curiosity

Is defined as the extent to which selecting the video is dependent on personal curiosity without having a previous interest in the topic or depending on the accidental discovery of useful or interesting information while searching for other information. This is different from Novelty where selecting the video is mainly based on it being new but also based on previous interest. Some responses mentioned Curiosity or being "curious" literally, for example, "curious to hear the song that everybody is singing at the University", "The title made me curious to know more details!"

In other cases, curiosity could be understood from the meaning and the context of the utterance, for example, "another video from the same YouTube channel. Just checking what kind of entertainment this channel provides", "the video subject interests me because he uploaded the video by mistake so I want see what he afraid of us to see". In some cases, the participants become curious about a video from the video thumbnail. For example, "the video picture was of man and woman with covered face, so I wanted to know why they covered their faces", "I saw a photo of a burger in black so I was curious to find out why this in black. Could be something dangerous I should avoid". Examples of discovering relevant information by chance include "I was seeking for Meghan Trainor's new single video clip 'No' but it was not released yet, then by chance from the resulting videos I found out that she had released a single 4 months ago which is 'Better when I am dancing'".

3. Habit

Is defined as the extent to which the participant is familiar with the video and watches it in a repetitive manner or as part of another habit. Responses classified under this code are similar to Familiarity responses in that participants selecting familiar videos. What distinguishes it from Familiarity is the mention of repetitively watching the video. For example, "I commonly watch this YouTuber and entertainment", "I just

love to listen this song over and over again" or watching the video as a part of another habit "Continued to fill out some applications and enjoyed listening to music as I worked", "To waste some time as I was eating breakfast I enjoy watching these videogame videos instead of watching cable tv" and "Go back to my routine and get ready to sleep". In the latter examples, YouTube is used as a background where another activity is going on. From the examples of Familiarity and Habit criteria, it is obvious how re-viewing or re-finding behaviours manifested in this research. When people keep watching the same videos because they are familiar with them or as a habit.

4. Time constraint

Defined as the extent to which time constraint is a factor in participant's judgement. This category is strongly related to Length criterion. Responses which simply mentioned the length (duration) of the video (too long or short) are coded as Length whereas responses which indicate that the participant's situation is a factor are counted under Time constraint. An example of these responses includes "I'd watch it till the end if I have time". So the reason behind the relevance judgement decision is more than just the length of the video, it is mainly based on the participant's situation and time that he could afford "I cannot afford to spend that much time", "It is too long, I will watch it at another time" and "It is too long and I should study".

Criteria related to the quality aspects of the video or the source providing the video

This group of criteria is mainly focussed on the video's perceived quality, either content quality, technical quality or the quality of the source providing the video.

1. Quality of source

Defined as the extent to which participant's judgement of the quality of the information is influenced by the source providing the video and whether the source

is well known or trusted. Examples of responses counted under this category include "It is also on one of the official YouTube channels linked to the programme, so it should be relevant", “I stopped watching because I can’t make judgment on fast food just from watching that. I need a more reliable source/s”. As mentioned before, when quality is assumed based on the familiarity with the source providing the video, the response is coded as Quality of Source not Familiarity, e.g. “I chose this video based on my familiarity with the channel, and it’s high production value, this, therefore, led me to believe this video would be like the others in terms of quality”.

When the video has many versions and the participant chooses the one provided by the official channel, these responses are coded under the Quality of Source rather than Version (see below), e.g. “It was not the music video therefore it would be the studio version with no interruptions”. In some cases, the source providing the information is a YouTuber or author, and a certain level of quality could be assumed based on those sources. For example, “the video was uploaded by a YouTuber that I know upload high-quality Video Game Music”.

2. Content quality

Is defined as the extent to which the video content is perceived to be of good quality. Responses related to the content rather than technical or source quality are classified under this code, e.g. “so I was interested to see if it was any good”, "because I don’t like the content (pictures without any description)" and "The recipe is too creamy and unhealthy".

3. Technical quality

Is defined as the extent to which image and sound are perceived to be of good quality, for example, “The image and voice quality in the video is quite acceptable”, and “I did not like the drawings of some of the characters When the video has many

versions and the participant chooses a high definition version, these responses are coded as Technical Quality rather than Version.

Criteria related to audio/visual features of the video

This group of criteria are more focussed on the aesthetic characteristics of the audio/visual content rather than the quality aspects of the video.

1. Cinematography

is defined as “the extent to which the video contained any specific film techniques the participant was interested in, such as camera movement, colour, editing, camera framing, special effects, and lighting” (Yang, 2005). An example of responses coded for Cinematography include “what’s more, despite being a vlog, it looks well shot (in terms of headroom, lighting, etc.)”, "interested to see how one of the iconic scenes of the series was filmed".

2. Visual appeal

is defined as the extent to which the thumbnail was appealing to the participant such as “the thumbnail was colourful”, "and the front image looks peaceful and appropriate and it works well for me" and "The picture of the recipes looks tasty and include desirable food item". Responses related to people presented in the thumbnail were not counted in this category, instead, such responses counted under the People in the video category.

3. Sound/voice

Is defined as the extent to which the participant likes the sound/voice content of the video. Examples of responses coded for this category include “I stopped the video early as I found the narrator’s voice irritating”, "because the singer’s voice is powerful and amazing" “have to watch the subtitles, because there is no talk only background

music” and "The tune is not my favourite, and it causes some distraction for me.". When the response mentioned the sound quality it is coded as Technical Quality.

Criteria related to the accessibility of the video

This group of criteria is related to how accessible the video is in terms of its cost, language or version.

1. Cost

Is defined as the extent to which some cost will be involved to obtain a video, for example, “to enjoy really good songs without having to pay for them”.

2. Language/subtitle

Is defined as the extent to which the language that was spoken in the video is understandable by the participant. For example, "I chose it because it is in English so I can understand what they said" and “it is in German”. If it was in a foreign language, whether there were subtitles shown in the video, e.g. “have subtitles”.

3. Version

Is defined as the extent to which different versions exist and judgements are based on the version of the video, e.g. “A different version of the art with the same actor and actress with the 1st video” and "I wanted to compare the live and studio version of the song".

Criteria related to other information within the environment

This group of criteria is related to the relationship between the information provided in the video to other information provided elsewhere.

1. **Availability**

Is defined as the extent to which a number of videos that cover the same topic are available and judgements are based on this aspect. This is not to be confused with Unusualness (below) which indicates that the video provides unusual information.

The information provided by the video based on Availability criterion might not be distinctive but the video was chosen because few videos are available about the topic, e.g. “few videos that cover desk decoration from this aspect”.

2. **Verification**

Is defined as the extent to which information provided in the video is consistent with or supported by other information or the extent to which the participant agrees with the information presented, such as “and I just couldn’t believe that a 3 years old would actually do that, so I had to check it out" and "this is impossible but I want to double check".

3. **Unusualness**

Is defined as the extent to which a video provides unique, weird or distinctive information comparing to other videos, for example, “This video was chosen based on the unique recipe it featured "and "There was a photo of a plant with teeth".

Criteria related to other people’s opinions or YouTube’s recommendations

This group of criteria related to being influenced by other people’s preferences or YouTube’s recommendations.

1. **Rank order**

Defined as the extent to which participant's decision to select a video is influenced by its position in the ranked list. For example "one of the first video to come up "and "Since there are many videos that upload the same song, I just click on the first one for convenience".

2. Popularity

Defined as the extent to which the video has a large number of views or likes. Examples "It has a large number of views", "it has 270 K likes, so it might be good" and "Most viewed music video so may be most popular best song by the band to start with and decide whether I like them". It could be noticed from some responses that participants predict some level of video's quality based on its popularity. For example, "it had over 2 million views so I could safely assume it was a reliable link".

3. Recommended video

Is defined as the extent to which a participant's judgement was influenced by recommendations provided by friends, YouTube, web pages or social media sites. Examples include "my friend suggested to see this video", "The suggested video to watch next", "someone recommended the video on Twitter".

Table 5.3 shows the frequency with which each criterion was used. Criteria related to the information content of the video were the most dominant category. Of the individual criteria, Topicality, Affectiveness and Recommended video were used more frequently in relevance judgements.

Category	**Criteria**	**Mentions**		**Participant**		**Mentions By Category**	
		#	**%**	**#**	**%**	**#**	**%**
Criteria related to information content of the video	Coverage	46	5.8	20	52.6	316	40.2
	Topicality	160	20.3	34	89.5		
	Recency	13	1.7	6	15.8		
	Genre	14	1.8	10	26.3		
	Length	33	4.2	15	39.5		

	People in the Video	50	6.4	21	55.3		
Criteria related to participant's previous experience and background	Background/ experience	10	1.3	7	18.4	100	12.7
	Novelty	36	4.6	13	34.2		
	Familiarity	54	6.9	20	52.6		
Criteria related to the participant's beliefs and preferences or situation	Affectiveness	88	11.2	27	71.1	128	16.3
	Serendipity/ Curiosity	21	2.7	9	23.7		
	Habit	14	1.8	6	15.8		
	Time Constraint	5	0.6	5	13.2		
Criteria related to the quality aspects of the video or the source providing the video	Quality of Source	36	4.6	13	34.2	67	8.5
	Content Quality	17	2.2	10	26.3		
	Technical Quality	14	1.8	7	18.4		
Criteria related to audio/visual features of the video	Cinematography	6	0.8	4	10.5	37	4.7
	Visual Appeal	11	1.4	5	13.2		
	Sound/Voice	20	2.5	10	26.3		
Criteria related to the accessibility of the video	Cost	1	0.1	1	2.6	11	1.4
	Language	7	0.9	4	10.5		
	Version	3	0.4	3	7.9		
Criteria related to other information within the environment	Availability	2	0.3	2	5.3	17	2.2
	Verification	8	1	5	13.2		
	Unusualness	7	0.9	5	13.2		
Criteria related to other people's opinions or YouTube's recommendations	Rank Order	17	2.2	9	23.7	111	14.1
	Popularity	22	2.8	9	23.7		
	Recommended Video	72	9.1	20	52.6		

Table 5.3 Number of mentions of relevance criteria

5.5.2 Gender and Culture Influence on Relevance Criteria

Gender and culture are studied as factors that influence the human behaviour in many research areas such as marketing and management. For example in a study which investigated the effect of gender on perceived online trust, functional magnetic resonance imaging used in a laboratory experiment to examine the brain activity of men and women when making trustworthiness decisions of eBay offers. The study reported gender differences as women had more brain areas activated than men (Riedl, Hubert, & Kenning, 2010). The study showed that similar product descriptions were treated differently between men and women, i.e. men and women were reading the descriptions differently. In another study which examined cultural differences in motivations for using social media sites, Kim, Sohn and Choi (2011)

surveyed college students from Korea and the USA to examine differences in motivations and patterns of using social media sites. The study revealed that the motives for using social media sites are similar between the two cultures, but the weight of these motives is different. Specifically, Korean participants found it more important to obtain social support from existing social relationships, while American participants emphasize more on entertainment as a motive to use social media sites. These findings provide practical implications to the social media websites marketers by advising them to pay attention to cultural differences and apply different strategies for different cultures to attract more users.

Li and Kirkup (2007) investigated both cultural and gender differences in use and attitudes of internet between China and the United Kingdom. Students from both countries were surveyed and significant differences were reported in Internet experience and usage between Chinese and British students. Gender differences were also found in both cultures and were higher in British participants. Men in both countries were more likely than women to use email or chat rooms and played more computer games. The study findings reflect on the implementation of information and communication technologies into education, suggesting to consider cultural and gender factors in this process.

In this study, the data was collected from two different countries (the United Kingdom and Saudi Arabia), and from both males and females. It is worth to examine whether users from different cultures or genders make different relevance decision and use different relevance criteria. If differences found, the video retrieval system should be sensitive to the main relevance criteria applied by certain gender or users from a certain culture.

Thus, relevance criteria applied by participants from both countries were qualitatively compared (as shown in Appendix A.12) to investigate whether cultural differences have an effect on relevance criteria selections. In general, there was no difference in the relevance criteria selections between the two groups.

In addition, as the gender distribution was not even in this study with more female participants, I compared the labels of relevance criteria mentioned between men and women participants. This was done for two reasons. First, to examine gender differences that might have implications for the video retrieval design. Second, to figure out whether more data need to be collected to balance the sample if gender has an effect on relevance criteria selection. Generally, the comparison showed no differences in applying relevance criteria between males and females (appendix A.12). This is in alignment with Reuter (2007) which found gender had only a slight effect on relevance criteria selection.

These findings indicate that applying relevance criteria in this study was consistent between cultures (KSA and the UK) and genders (men and women). Although the sample includes fewer men than women, the majority of the criteria mentioned by female participants are also covered by male participants (except Cost and Language criteria). The same is applied to the two different cultures, all the relevance criteria (except Cost which only mentioned once) mentioned by participants from the United Kingdom were also mentioned by the participants from Saudi Arabia. Different groups of gender or culture might still use relevance criteria differently in terms of the frequency of mentioning a relevance criterion. Quantitative comparisons of the number of mentions of each relevance criterion between the two groups could not be applied in this study because the participants sample was not balanced between males and females. However, the key result of these comparisons is having the same set of relevance criteria between genders and the two cultures.

5.5.3 Topics and Motivations

The participants provided various search topics when answering the “What are you looking for?” diary question. In many cases, the related “Why did you start this search?” question, helped in clarifying the topic by giving more details of the context and motivation of the search. For example, the reason for searching the topic “derma

roller" was clarified in the motivation field by this statement "to know about it before buy it".

In this study data, 62% of leisure searches began with a specific information need (topic) where the participants are looking for videos about a topic in mind (e.g. movie trailer). The participant might be looking for videos very specific to the topic (e.g. specific movie name) or more generally videos from the same type or genre (e.g. movie trailer 2018). The main topics mentioned in the diaries are summarized in Table 5.4.

The topics mentioned in this study overlap with the topics mentioned in previous studies of video retrieval. For example, Cunningham and Nichols (2008) categorized the video information needs in their study into: music, humour, movies, TV, computers, cars, sport and other. Yeh's study of casual-leisure information behaviours for watching videos online found that participants' videos were about the following topics: movies, dramas and entertainment television shows (Yeh, 2016).

Topics	**Examples**	**Count of mentions**
Music	"Adele hello" "pop music"	36
Beauty	"Hairstyle tutorial" "how to take care of curly Hair"	18
TV or YouTube shows	"best of the voice 2015" "The Ellen Show"	14
Movie trailers	"I know what you did last summer video" "movie snippets"	14
Celebrities	"Antony Robbins video"	7
Video games	"looking for a new game to play"	7
Drama series	"tv series"	7
Comedy clips	"funny videos"	6
Cooking	"Sweet potato recipe"	6
Animals and plants	"strange animal partnerships"	4
Decoration	"I was, in this case, in search of decoration ideas"	3
Sports	"to look at sports and people"	3
Travel and tourism	"Okun island"	3
Motivational speeches	"motivation video"	3

other	"Videos about US presidential race" "magic tricks"	13

Table 5.4 Participants' topics for leisure searches

The remaining search sessions lack the mention of specific topics and were triggered by various motivations. The motivations of leisure searches are summarized in Table 5.5.

Motivations	**Examples**	**Count of mentions**
Pass Time	"Getting something to watch during waiting time" "Kill the time"	44
Change Mood	"change my mood" "I was feeling bored and I wanted something to cheer me up"	27
Relaxation and Refreshment	"wanted to relax after a long day and before going to evening work" "Take a break and refresh my mind while working on a website"	19
Learning	"I want to learn how to make a natural look for everyday" "wanted to learn a new dish for dinner"	16
Background or Distraction	"my usual background when I am doing body exercise" "to block out building noises in my house which is having a loft extension"	13
Recommendations	"a friend recommended the band to me so I checked them out" "I would like to watch the trailer of the latest movies, because posters of some of new movies look quite interesting"	11
Entertainment	"for some laughs" "I went on YouTube purely seeking entertainment."	10
Other	"our teacher showed it to us and I want to show it to my sisters" "having trouble sleeping" "my daily habit of checking my subscriptions"	13

Table 5.5 Participants' motivations for leisure searches

The motivations for searching videos at leisure time mentioned in this study are in line with the previous studies. Cunningham & Nichols (2008) classified the

motivations for searching videos into eight categories: mental status, visual, audio, learning, social, mainstream media, temporal and other. Despite the differences in the labels, “mental status” description (which is “explicit reference to subject’s emotional state/mood”) and the examples provided matched examples classified under “pass time”, “change mood” and “entertainment” in this study. “Learning” category is also common in both studies. In addition, “social” motivation includes examples of users who look for videos because it was recommended or to share it with others. In this study, such responses were classified as “recommendations” and “other”, respectively.

The motivations that stimulated the participants to initiate their searches in this diary study are also in agreement with the findings of Yeh (2016) who divided participants’ search sessions into three stages: pre-viewing, viewing and post-viewing. Yeh differentiated between two ways the participants approach the information of the video in the pre-viewing phase: actively by searching for information or passively by being triggered by other information. Accordingly, motivations could be divided to motivations for active search (passing time, search out of boredom) and motivations of passive search (recommendations by peers, information encountering).

5.5.4 Criteria Used at Different Search Stages

During the relevance judgment process, participants apply relevance criteria at different stages of the search. Some criteria are applied before watching the video when the participant predicts the video is going to be relevant. Others were applied after watching the full video and many could be applied at both stages. In many cases, it is not possible to identify the stage of the search from the utterances in the diaries.

Relevance criteria exclusively used before watching the video include criteria related to other people’s opinions or YouTube’s recommendations (Rank Order, Popularity and Recommended Video), Curiosity, Cost, Visual Appeal and Recency. All the mentions of these criteria indicated a predictive relevance judgment decision.

Examples include "'it has 270K likes, so it might be good", "'Looks like quite new". However, after watching the whole video, other criteria could be applied. For example, participant could predict the video to be relevant based on YouTube recommendation but then found it not useful because of the lack of Topicality, Content Quality or Affectiveness. Usually these criteria are mentioned with the criteria related to other people's opinion and YouTube's recommendation.

On the other hand, only criteria such as Sound/Voice criterion could not be applied unless the participant has watched the video. For example, "The music is too loud for me; therefore, I have to stop playing it".

The majority of the remaining relevance criteria were mentioned in both stages. In some cases, the stage of the search could not be extracted from the participant's utterance. As an example, the Coverage criterion could be applied at the predictive stage (before watching the whole video) e.g. "'from the title it is clear it is going to be useful and brief", or after watching the whole video "'he's giving too much details" or the stage of the search could not be recognized e.g. "'I just need to see some brief of the game to get the impression". The evolution of relevance criteria selection among search stages is not investigated further in this study but is investigated in the dynamic use of relevance criteria study (Chapter 6).

5.5.5 Positive versus Negative Mentions of Relevance Criteria

Relevance criteria could have a positive or a negative contribution to the relevance judgment decision. In other words, they could contribute to judging videos as relevant or judging them as non-relevant. By observing the relevance criteria mentioned in this diary study, criteria could be divided into three groups: criteria that are mentioned both negatively and positively, criteria that are only mentioned positively and criteria that are always mentioned negatively as shown in Table 5.6.

All the criteria applied in the predictive stage of the search process have only positive mentions. However, I should acknowledge that there might be other criteria applied

negatively in the predictive stage, but the participants did not report them. Besides the criteria applied in the predictive stage, other criteria such as Familiarity have only positive mentions. This is in agreement with findings of (Savolainen & Kari, 2006) who reported that Familiarity is frequently used to judge hyperlinks and pages as relevant.

Criteria mentioned positively and negatively	**Criteria always mentioned positively**	**Criteria always mentioned negatively**
coverage	popularity	availability
topicality	rank order	time constraint
length	recommended video	
background/experience	version	
novelty	visual appeal	
affectiveness	cinematography	
technical quality	habit	
content quality	serendipity/curiosity	
quality of source	familiarity	
sound/voice	people in the video	
language	genre	
verification	recency	
unusualness		

Table 5.6 Positive versus negative mentions of relevance criteria

5.5.6 Information Elements used when Applying Criteria

Some previous relevance criteria studies mixed between relevance criteria and information elements, Ruthven and Kelly (2012) stated that information elements are used as clues that aid the user in applying his criteria in order to make a relevance decision. While the main focus of this study is to investigate relevance criteria applied in leisure context, the study can also shed the light briefly on those information elements which used by participants while making their relevance judgment decisions. Table 5.7 illustrates the range of criteria that applied on various information elements.

Element Mentioned	**Criteria Applied**
Thumbnail	Visual Appeal, Affectiveness, Cinematography, Coverage, Curiosity, People in the video, Content Quality,

	Technical Quality, Topicality, Unusualness, Verification.
Title	Topicality, Affectiveness, Coverage, Curiosity, People in the video
YouTube channel name	Quality of Source, Familiarity
View Count	Popularity
Date of upload	Recency
Video Duration	Length
Video position in the result list	Rank Order

Table 5.7 Information Elements

The most dominant element which was used by 11 relevance criteria is the thumbnail of the video. Examples, where thumbnail could be useful when the participant is judging the video, include Cinematography criterion “a well-framed picture of food” or Curiosity “the video’s picture was of a man choking himself. That gets me interested to see why he was doing that”. Visual appeal criterion is totally depending on the thumbnail “and the picture of the video is attractive”. Technical quality could also be assessed or predicted from the thumbnail e.g. “I picked this video as it looked of a high production quality (assumed through the thumbnail)”.

The title of the video was mentioned when applying five of the relevance criteria. For example, Affectivness criterion could be applied based on the title of the video “The title said it 'so funny'”. Another example is Curiosity, a participant could be curious about the video because of its title “the title of this Ted talk about Arab so I have been took by curiosity”. Topicality criteria usually applied based on the title of the video e.g. “The title matched my search”. As the information elements used when applying relevance criteria were not the focus of this study, participants were not asked directly about them. Hence, it is not always possible to identify what element helped the participant in making this assessment. Some responses mentioned the element clearly others not.

5.6 Limitations

The decision to follow a naturalistic approach reduced the level of control on the study. First, the diary method depends on the participant's commitment to keep filling out the diary, thus the data collected from each participant did not necessarily captured all the search sessions conducted during the one-week diary study. Moreover, the participants might be selective in recording their sessions and avoid reporting sessions which seems personal or embarrassing. Second, the instructions provided to the participants stated that participants should record their leisure sessions while or soon after they conduct the search to avoid memory lapses, but this is not guaranteed to happen and the participants might forget to do so. Third, a limitation of the methodology can be found in the subjective nature of the analysis of qualitative data. The content analysis was mainly based on what the participants say in their diaries. The participants' ability to express their relevance criteria applied in the relevance judgement process varies from one to another and the information provided by the participants was on varying levels of details. In case of brief diaries, my interpretations of the participant's inputs might not totally in line with what the participant intended to say. A further limitation can be found in the participants' sample used in this study, all the participants were university or college students.

5.7 Chapter Summary and Implication for the Second Study

This chapter presented the study which was undertaken to investigate relevance criteria when searching videos in leisure context. Previous relevance criteria studies mainly focused on text retrieval in academic or work-related context, thus, a study which investigates video relevance criteria in leisure context was needed. To address this goal, a naturalistic diary study was applied and 30 participants completed the diaries. The analysis revealed 28 video relevance criteria in the leisure context. Criteria related to the information content of the video were the most dominant category. Of the individual criteria, Topicality, Affectiveness and Recommended video were used more frequently in relevance judgements.

Although the diary study revealed the list of relevance criteria applied when searching for videos in leisure contexts, the study did not investigate the use of relevance criteria at different stages of the search process, in other words, the dynamic use of relevance criteria. As mentioned in Chapter 2 Section 2.1.3.2, dynamic relevance criteria were discussed in relevance criteria studies conducted in academic contexts by adopting the stages mentioned in various information seeking frameworks. Thus, an open question is to investigate how relevance criteria choices might change at different stages of a leisure search.

Chapter 6

Dynamic Use of Relevance Criteria

Chapter 5 provided basic investigation of relevance criteria people applied when searching for videos in leisure contexts. The method used was a naturalistic diary study and the findings reported the main relevance criteria applied in this context. As relevance criteria are known to be dynamic, another study is needed to examine the dynamic aspects of relevance criteria at different stages of video/ leisure search. This chapter reports on the findings of the second study of this research. The purpose of this study is to examine the differences in relevance criteria at the stages of selecting and viewing videos for leisure. The chapter begins with a brief overview of the study then introduces the study participant's sample. Section 6.3 provides the detailed procedure of the study followed by the data analysis in Section 6.4. Then the findings of the study are presented in Section 6.5. The chapter concludes with the limitations of the study and chapter summary.

6.1 Overview

Previous research (Bateman, 1997; Schamber et al., 1990) has shown that relevance is dynamic and that users' relevance judgments can change over time. This evolution in relevance judgment is a reflection of the evolution in relevance criteria choices. These previous studies have been focused on academic contexts with the result that little is known about the dynamic aspects of relevance criteria used in leisure contexts, specifically for video content. Understanding how relevance criteria selections evolve as the search progresses and what criteria are more important at

specific stages of the search will provide a deeper understanding of the dynamic aspects of relevance criteria and will have implications on the design of IR systems. IR systems should be more adaptive to the change in users' preferences of relevance criteria during the search and support users with useful information needed for relevance judgment decisions as they progress in their search. Relevance judgments of video content are more complicated than text or images as the users' needs are diverse and videos could be judged based on visual and audio features besides textual ones (Yang, 2005). Therefore, this study attempts to fill in this gap by investigating how participants might change their video relevance criteria selections at different stages of leisure searches. A secondary aim of the study is to examine the effects of changing the methods of data collection on the findings. In the first study, data was collected in naturalistic settings using diaries, whereas this study uses a more controlled approach: recorded search sessions followed by interviews. The two methods might have different results because of the differences in the study settings, thus relevance criteria resulted in this present study will be compared to the findings of the first study (investigating relevance criteria for videos in leisure context). The study aims to address research questions RQ3 and RQ4. Full discussion and justification of the method selection and pilot studies preceded the main study were provided in Chapter 3.

6.2 Participants

Twenty-four YouTube users participated in the study. None of them had participated in the diary study. Of the participants, 13 were males and 11 were females. The range of ages is between 19 and 58 with an average age of 27.5. Among the participants, 12 were undergraduate students, six master students, four university staff one college student and one PhD student. They were studying different disciplines at the University of Strathclyde. Around half of the participants reported using YouTube several times a day, ten of them search from one to three times a week and only three reported that they search one or twice a month. The full demographic data of the participants of this study is shown in Table 6.1. Participants were recruited

through flyers distributed in different places in the campus and £5 were offered as compensation for their effort.

P#	Gender	Age	Course of study	Level	Frequency of use
P1	F	42	Medication and conflict resolve	master	Several times a day
P2	F	31	Information and library study	master	Several times a day
P3	M	50	PG Cert Teaching and Learning	university staff	3 or more times a week
P4	M	24	Sustainable engineering	master	3 or more times a week
P5	F	23	Investment and finance	master	1-2 times a week
P6	M	22	Business analysis and consulting	master	Several times a day
P7	F	21	mathematics	undergraduate	3 or more times a week
P8	F	19	Forensic and Analytical Chemistry	undergraduate	Several times a day
P9	F	20	history	undergraduate	1-2 times a month
P10	F	58	-	university staff	1-2 times a month
P11	M	21	pharmacy	master	3 or more times a week
P12	M	21	Computer science	undergraduate	3 or more times a week
P13	M	33	Mathematics ad	undergraduate	Several times a day
P14	M	30	Energy Policy	university staff	Several times a day
P15	M	35	Information science	PhD	Several times a day
P16	M	20	History and Politics	undergraduate	Several times a day
P17	F	26	Digital media	College student	Several times a day
P18	M	21	Aero Mechanical Engineering	undergraduate	Several times a day
P19	M	20	Aero-Mechanical Engineering	undergraduate	1-2 times a week
P20	F	19	mathematics	undergraduate	1-2 times a month
P21	F	21	psychology and English	undergraduate	1-2 times a week
P22	F	21	History and French	undergraduate	1-2 times a week
P23	M	19	Civil Engineering	undergraduate	Several times a day
P24	M	42	-	university staff	3 or more times a week

Table 6.1 Participants' demographic information

6.3 Procedure

Participants were invited to a private room in the university. The study purpose and procedure was explained verbally and as a handout provided to the participant. Then, the participant signed a consent form (Appendix B.4) in order to indicate agreement to participate in the study. Each participant was asked to fill a form prior to start the search session that collected demographic information and the frequency of searching YouTube for leisure (Appendix B.6). To prepare the participant for the leisure search session, a short chat about the participant's motivations for viewing videos and the types of video that the participant would normally watch at free time proceeded the actual search session. The questions asked in this pre-search interview are provided in Appendix B.7. Then the participant was given 20 minutes to search YouTube as normal. The participant was informed that the search topic should be personal and non-work or study related. There was no restriction on how the participant initiates the search, e.g. starting by typing a query, YouTube recommendations, or subscription channels. Appendix B.5 provides the search scenario given to the participants prior to their searches. All search sessions were recorded using Camtasia screen recording tool. On completion of the search session, I returned to the room and played back the session. A semi-structured interview was conducted in which the participant was able to watch back the search session and describe the reasons behind their relevance judgment decisions on each video for both selection and viewing stage. Selection stage is the stage where participants are evaluating the videos in the search result list or browsing videos in the home page of YouTube or a specific channel. For each video the participant decided to click on, she was asked what attracted her to click on this video and what made her predicted that the video would be relevant? The viewing stage represents the actual viewing of the video. For each video the participant viewed, he was asked what he thinks about the video during watching it.

Following this approach, I believe that I avoided the distraction that might be caused if I asked questions while the participant is watching his videos. All interviews were audio recorded and transcribed.

6.4 Data Analysis

Data analysis began with careful reading of the transcripts, noting utterances that reflected relevance criteria mentions. The stages of the search process applied in this study followed Yeh's (2016) framework of casual-leisure video viewing processes and information behaviours. The framework divided the process of viewing videos online into three phases according to time: pre-viewing, viewing, and post-viewing (Figure 6.1). In this study, I refer to the pre-viewing stage as the selection stage. As there were no relevance judgments after viewing the videos, the analysis only focused on the selection and viewing stages. This framework was chosen because it shares similarities with this study in terms of the context (casual leisure video viewing), but the focus is different. Yeh's study investigated the motivations that trigger casual leisure video search and the information behaviour activities while viewing videos. In contrast, this study focuses on the differences in relevance criteria applied between stages of the search. Other models such as Kuhlthau (1993) and Ellis (1993) were used in dynamic relevance criteria studies that investigate academic context, but they are not applicable in the leisure search context. These models identified several stages that users go through and have some assumptions that not always held in the leisure contexts search. Kuhlthau's model, for example, has six stages, the first is task initiation in which the user recognizes the lack of knowledge and feel uncertain. Lack of knowledge and uncertainty is not always the initial stage of leisure searches, people might initiate their searches triggered by other hedonic motivations. Similar to Kuhlthau, Ellis's model identified six stages of the search process and these stages have an academic focus. For example, the *chaning stage* which indicates the user behaviour when following the chains of citations from one source to the other.

Each utterance was coded based on the coding scheme of the first study (investigating video relevance criteria study) with the possibility of adding new codes when needed. Each mention of relevance criteria was assigned criteria code and a code to note which search stage: selection or viewing. Video recordings were viewed to get a better understanding of the session but the main analysis depends on the transcripts.

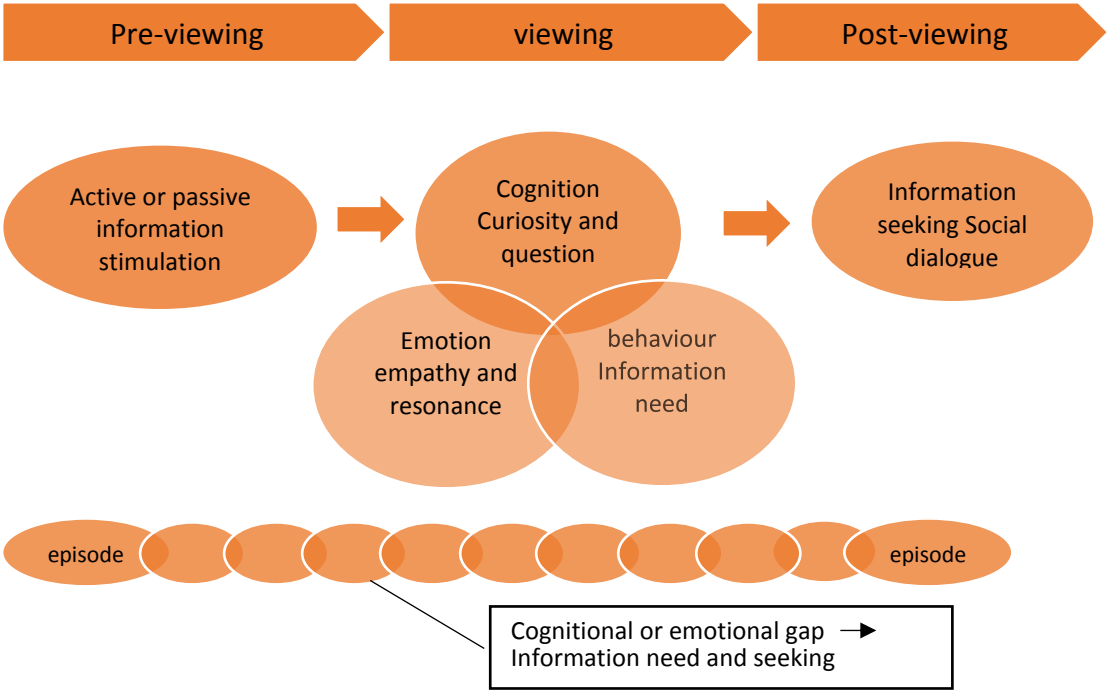

Figure 6.1 Yeh's framework of video viewing processes and information activities (Yeh, 2016)

To investigate the differences in applying relevance criteria between the selection and viewing stages, count of number of mentions of each criterion are provided for each stage (Table 6.4). Chi-squared test was applied to examine whether the variance in number of mentions of each criterion between the two stages is statistically significant. Thus, Chi-squared test applied to investigate whether a relationship exists between the stage of the search and relevance criteria applied and to ensure that the relationship is a valid relation and not due to chance (Little, 2013). Chi-squared test

is selected as it is a nonparametric test suitable for categorical data and it is “the most popular significance test used to analyse frequency counts”(Lazar et al. 2017, p.94).

6.5 Findings

In total, 818 mentions of relevance criteria were revealed by the interviews’ transcript analysis. The total number of videos watched was 165 videos with an average of seven videos per participant (min 3, max 16, SD 3). Participants searched for various topics including: songs and music, TV or YouTube shows, movie trailers, celebrities, video games, comedy clips, animals’ videos, sports, travel and tourism, motivational speeches and news. As we can see, the topics mentioned in this study are similar to the topics mentioned in investigating relevance criteria for videos in leisure context study Chapter 5 Section 5.5.3

The findings are structured as follow: first the differences in relevance criteria between selection and viewing search stages were investigated. Then, the findings from this study are compared to the findings of the previous study (Chapter 5). Finally, further investigation of the change in relevance criteria between the start and end of the sessions is applied.

6.5.1 Relevance criteria at different search stages

Following Yeh’s (2016) stages of casual leisure searching for videos online, this section will examine participants’ relevance criteria at the selection and viewing stages of the search process. Then, in order to answer the third research question, what is the difference in employing relevance criteria between selecting and viewing stages of video/leisure contexts searches? Differences between the two stages in terms of the mentions of relevance criteria will be discussed in Section 6.5.2.

6.5.1.1 Selection stage

The selection stage precedes the actual viewing of the video where participants select videos to watch from YouTube homepage, specific channel page or from the search result list of their queries. Search sessions started by either searching for a specific topic or browsing videos on the home page. Only few participants logged in to their accounts and browsed their subscriptions. The data showed that participants mainly predict the relevance of a video based on Topicality which was the most dominant criterion in this stage, accounting for nearly one-fifth of the mentions. Examples of Topicality responses include: “Even the name gives it away, which is one of the reasons I clicked on it after seeing that James Corden was on it” p22, “I searched for Air Force One the aircraft. And chose this video because of the title says the inside and then the secrets of presidential travel” p13 and “Because I'm looking for a job, Then the topic is really relevant to what I'm doing now” p5.

The second most mentioned criterion in the selection phase is Familiarity. For example “and also with something like that, I’ve seen shows like that before. Because I’d seen videos like this before, like on UK TV shows. And it’s always been funny in the past” p22 and “So I type in a game that I’m already familiar with” p23.

Topicality and Familiarity of the video are predicted from the title or the thumbnail. Sometimes participants recognize familiarity with the source from the channel name as well, for example, “also I recognize the uploader as well” p16. Familiar videos are preferred in some cases to guarantee enjoyment and to save participant's time:

> "I do not click on videos from people that I do not know. If its news, or world events, then that’s okay, watching Sky News or CNN. But that’s just on a kind of event to event news basis. Usually I stick with what I know, just because I don’t want to waste my time" p17

This interesting behaviour might vary among different participants and could be further investigated in future work to examine what user variables (such as

personality) could affect relevance criteria choices. Participants also followed recommendations provided by friends, YouTube or social media sites. The Recommended video criterion acquired 9.8% of the total mentions in this stage. Examples: “Because it was recommended in this section. So I just clicked on it” P23 or “I think I clicked on Shakira because I follow her on Instagram. And I’d seen pictures of this music video” p21. Participants may select videos to watch based on their Novelty “but I hadn’t watched it, so I decided I would watch it just now” p17, “so I just started looking through videos and I found a new Nietzsche video, which I was like, ‘I haven’t seen this’” p12.

People going to appear in the video was another reason for selecting videos. For example “It's only the people he has in his car that I like that I'll watch. He has other people in his car that I won't watch” p1. Another example is:

> “Who is in the pictures and who was in it, and also I really like Ed Sheeran. So I thought 'oh, that will be funny', so I watched that. Because it was featuring someone that I actually knew and that someone that I liked” p22.

Novelty and People appeared in the video each acquired 8.8% of the overall mentions of relevance criteria. During the selection stage, participants also give attention to the source providing the video and its quality. For example, “It’s Warner Bros. which is a move company, so you know it’s real, it’s not a fake trailer or anything” p23, “oh this one, because it’s an official news video, I decided to give that a bit more time. Just to see a more professional angle” p17. There are some criteria that exclusively mentioned in the selection stage. For example, the appealing of the thumbnail of the video (Visual appeal) “and the picture made it look really funny because she was pulling a face or something”p22, “But it was the thumbnail that attracted me because it is the same crazy hair that Trump has. Yes. Which is this alternate right-wing politician in Holland going for the...” p15. Other criteria that only mentioned in this stage include: Rank order, Recommended videos and Serendipity/Curiosity. The full

list of relevance criteria in the selection stage ordered by their number of mentions is provided in Table 6.2.

Criteria	#	%	Criteria	#	%
topicality	75	18.2	rank order	8	1.9
familiarity	45	10.9	version	8	1.9
recommended video	40	9.7	serendipity/curiosity	7	1.7
people in the video	37	9	content quality	4	1
novelty	35	8.5	genre	4	1
quality of source	29	7.1	language	3	0.7
visual appeal	21	5.1	habit	2	0.5
popularity	19	4.6	technical quality	1	0.2
length	16	3.9	cinematography	1	0.2
recency	15	3.6	sound /voice	1	0.2
background experience or personal memories	14	3.4	unusualness	1	0.2
coverage	13	3.2	verification	1	0.2
affectiveness	10	2.4	time constraint	1	0.2

Table 6.2 Mentions of relevance criteria in the selection phase

6.5.1.2 Viewing stage

During viewing stage, Affectiveness was the most dominant criterion with approximately one-fifth of the overall mentions of relevance criteria in this stage. The effect that the video selected made on the participant's feeling in a negative or positive way is important to the participants at this stage. For example, a participant might express positive effect of the video by comments such as "I find it interesting. It's emotionally engaging for me" p3, "he has a particular humour, let's say, which I like, which is very ironic and sarcastic" p14 or negatively "And then I got bored with that" P3, "I'll turn it off in a second, because I got bored. Because it wasn't what I expected, and I didn't find it as amusing" p22.

In the viewing stage, Topicality continued to be an important criterion with 10.5% of mentions. Topicality was mentioned to express the matching between the video's information and the participant's search topic, e.g. "And in the end it wasn't that much I have to say. It was just funny jokes but not really directed to the topic I wanted" p14. Or a topic that participants want to know more about "Because this

video gives you information on a game that's coming out at the end of this year" p23. Topicality could also be describing the content or the aboutness of the video "And what it's actually about is about Coventry in 1987" p3, "So this is a video about the annoying people you find in Glasgow parks. People barbequing and playing Frisbee and being loud with music but it does it in a funny way" p2.

Besides being on topic, participants pay attention to the Content Quality of the video. For example, a participant might select a video on a specific topic but then get annoyed by the poor quality of the information provided in the video, e.g. "It was too informal, too unprofessional. It was kind of just mashed up. There was no proper information, it was just clips" p17. Another example of Content Quality:

> "This one is better because he's describing what's going to be in it. It's more kind of factual, actually explaining it. Whereas the one before, he was just playing it, but this one he's explaining 'oh, the new map's really good, the new gun's really, really good' " p6.

Whilst Novelty was important in selecting videos, it could be a crucial criterion in discarding a video after starting to watch it, for example, "For this one, it wasn't even a very funny one, so as soon as I recognized that I had definitely seen it before, I clicked away" p2. Novelty also applied for positive relevance decision "It's something I haven't seen before, so that definitely piqued my attention" p17.

During viewing stage, Layout/presentation of the information as appeared in the video was mentioned as a criterion for relevance judgment. For examples,

> "Apart from the video, it's just the layout. It says that it is inside the aircraft, so it shows you where the aircraft is based. And then actually how to get into the aircraft, and then it shows you the different compartments" p13.

Another example: “I like the topic and the way it’s delivered” p2. This criterion is exclusively mentioned in the viewing stage.

Around 6% of the mentions of relevance criteria during viewing stage related to the Coverage aspects of the information provided in the video, for examples: “Yeah, yeah, it’s informal. But it’s very informative, and it cuts to the point. And it shows you what you want to see. I wasn’t left thinking ‘what if…what did I not see?’” p17, “In-depth, but like not exaggerated, yes. ‘This is the office, it is connected to that, that, that’. Taking you through the whole aircraft, but in a very reasonable time” p13. Participants mentioned a wide variety of additional criteria in this stage, such as Technical Quality and Cinematography. Table 6.3 provides the full list of the mentions of relevance criteria in viewing stage.

Criteria	**#**	**%**	**Criteria**	**#**	**%**
affectiveness	88	21.6	unusualness	14	3.4
topicality	42	10.3	quality of source	7	1.7
content quality	40	9.8	language	7	1.7
novelty	33	8.1	recency	6	1.5
layout	25	6.1	version	5	1.2
coverage	24	5.9	verification	4	1
background experience or personal memories	20	4.9	familiarity	3	0.7
technical quality	18	4.4	genre	3	0.7
length	18	4.4	habit	2	0.5
cinematography	16	3.9	popularity	1	0.2
people in the video	15	3.7	time constraint	1	0.2
sound /voice	15	3.7			

Table 6.3 Mentions of relevance criteria in the viewing phase

6.5.2 Differences between Selection and Viewing Phases

This section will show the changes in criteria selections between selection and viewing stages. Table 6.4 specifies the differences in the number of mentions of each relevance criterion between the two search stages.

Relevance Criteria	Selecting #	Selecting %	Viewing #	Viewing %	Total
Criteria related to the information content of the video					
coverage	13	3.2	24	5.9	37
topicality	75	18.2	42	10.3	117
recency	15	3.6	6	1.5	21
genre	4	1	3	0.7	7
length	16	3.9	18	4.4	34
people in the video	37	9	15	3.7	52
Criteria related to the participant's previous experience and background					
background experience or personal memories	14	3.4	20	4.9	34
novelty	35	8.5	33	8.1	68
familiarity	45	10.9	3	0.7	48
Criteria related to the participant's beliefs and preferences					
affectiveness	10	2.4	88	21.6	98
serendipity/ curiosity	7	1.7	0	0	7
habit	2	0.5	2	0.5	4
time constraint	1	0.2	1	0.2	2
Criteria related to the quality aspects of the video or the source providing the video					
quality of source	29	7.1	7	1.7	36
content quality	4	1	40	9.8	44
technical quality	1	0.2	18	4.4	19
layout	0	0	25	6.1	25
Criteria related to audio/visual features of the video					
cinematography	1	0.2	16	3.9	17
visual appeal	21	5.1	0	0	21
sound/voice	1	0.2	15	3.7	16
Criteria related to the accessibility of the video					
language/subtitle	3	0.7	7	1.7	10
version	8	1.9	5	1.2	13
Criteria related to other information within the environment					
verification	1	0.2	4	1	5
unusualness	1	0.2	14	3.4	15
Criteria related to other people's opinions or you Tube's recommendations					
rank order	8	1.9	0	0	8

popularity	19	4.6	1	0.2	20
recommended video	40	9.7	0	0	40
Total	**411**	**50.2**	**407**	**49.8**	**818**

Table 6.4 Comparisons of the mentions of relevance criteria at the selecting and viewing phases

Chi-squared test was conducted to examine the differences in relevance criteria mentions between selection and viewing stages. The null hypothesis to be tested is "there is no significant difference in using relevance criteria between the selection and viewing stages". Similar to previous studies (Maglaughlin & Sonnenwald, 2002; Savolainen & Kari, 2006) share of relevance criteria are uneven with some criteria being crucial while others marginal. Marginal criteria with low-frequency count (less than ten) were not included in the test (Verification, Time constraint, Serendipity, Rank order, Habit and Genre).

Taken as a whole, the test result revealed a highly significant difference in applying relevance criteria between selecting and viewing stages, $\chi 2\ (20) = 325.103$, $p < .001$ and the null hypothesis was rejected. Some criteria are more important at the selection stage while others have more mentions in the viewing stage. A follow-up post hoc test was performed following the 'calculating residuals' approach to identify the criteria which contribute to the significant variance between the two search phases (Field, 2013; Sharpe, 2015). A Bonferroni correction is suggested when the number of comparisons is high (comparing selecting and viewing stages for 21 relevance criteria) to avoid Type I error (Sharpe, 2015; Macdonald & Gardner, 2000).

A Bonferroni correction was conducted and the corrected alpha was $\alpha = .002$ Table 6.5 and 6.6 show the criteria that were statistically significant in the selecting and viewing stages respectively. If the Bonferroni correction was not applied, the number of mentions of an additional criterion (Recency) would also be significantly varied between the two stages.

Criteria	**χ^2**	**# of mentions**
recommended video	43.16	40

familiarity	40.20	45
visual appeal	22.09	21
popularity	17.06	19
quality of source	14.59	29
topicality	11.83	75
people in the video	10.50	37

Table 6.5 Statistically significant criteria in the selection stage

Criteria	**χ^2**	**# of mentions**
affectiveness	68.89	88
content quality	30.36	40
layout	25.20	25
technical quality	15.21	18
cinematography	13.18	16
sound	12.18	15
unusualness	11.16	14

Table 6.6 Statistically significant criteria in the viewing stage

Throughout the stages, some criteria (such as Novelty and Length) remain steady, others changes slightly but did not contribute to the significant difference between the stages. Moving from the selection to the viewing phase, the number of mentions of Familiarity, Topicality, Source Quality and Popularity significantly dropped. Criteria that only mentioned in the selection phase are Recommended Video and Visual Appeal. Conversely, criteria which show a significant increase in the number of mentions are Affectiveness, Content Quality, Technical Quality, Cinematography, Sound and Unusualness. Layout criterion was only mentioned in the viewing stage. It is expected that some of the criteria such as (Technical Quality and Cinematography) will have more mentions in the viewing stage as the participant needs to examine the full video to judge it based on these criteria. The participant could predict relevance based on these criteria at the selection stage but will not be determined until the viewing of the video itself. On the other hand, other criteria such as Source Quality and Popularity mentioned more in the selection stage as participants predict relevance based on them.

6.5.3 Comparing the Findings with the previous Diary Study

One of the few works that compare diary as a data collection method to other methods was Greenberg et al., (2005). The authors conducted a study to compare diary to survey in terms of measuring individual's internet and traditional media use. The findings showed that participants reported higher estimation of internet and traditional media use in the survey, but both methods are correlated with each other. Other methodological comparisons studies exist in different fields such as economics and anthropology, (Brzozowski, Crossley, & Winter, 2017; Paolisso & Hames, 2010) however, the research areas are too different to inform this research.

To answer the fourth research question, does the diary method provide different findings from recorded search sessions with interviews? The number of mentions of each criterion between the two studies was compared. Preliminary analysis showed results which support the previous diary study's findings. Table 6.7 contains the full comparison between the two studies. Both studies showed that criteria related to the information content of the video were the most dominant category. The percentage of the total mentions of this group of criteria was 32.8% in this study and 40.2% in the previous diary study. Of the individual criteria, Topicality and Affectiveness were used more frequently in relevance judgments. Topicality is the most common criteria in both studies followed by Affectivness. Marginal groups of criteria have also been found less influential in this study compared to the diary study. Specifically, criteria related to the accessibility of the video and criteria related to other information within the environment were the two lowest mentioned groups of criteria. The number of mentions of Recommended Video criterion had dropped in this study compared to the diary study, but I should acknowledge that in many cases even if this criterion was not mentioned explicitly in the interviews with the same percentage of the diary, it was used a lot implicitly as participants in the majority of cases were selecting their videos from the recommendation list by YouTube. Layout/presentation label was added to the coding scheme which is defined as the extent to which presentation, delivery and clarity of the information are factors in

participant's relevance judgment. Examples include, "I like the topic, and the way it's delivered" p17, "And then this is the guy doing the review, and he's got a much more interesting approach and presentation. He's actually presenting the thing to you" p3.

Criteria	**Interview mentions**		**Interview Participants**		**Diary mentions**		**Diary participants**	
	#	**%**	**#**	**%**	**#**	**%**	**#**	**%**
Criteria related to the information content of the video	**268**	**32.8**			**316**	**40.2**		
coverage	37	4.5	11	45.8	46	5.8	20	52.6
topicality	117	14.3	23	95.8	160	20.3	34	89.5
recency	21	2.6	11	45.8	13	1.7	6	15.8
genre	7	0.9	5	20.8	14	1.8	10	26.3
length	34	4.2	14	58.3	33	4.2	15	39.5
people in the video	52	6.4	20	83.3	50	6.4	21	55.3
Criteria related to the participant's previous experience and background	**150**	**18.3**			**100**	**12.7**		
background experience or personal memories	34	4.2	14	58.3	10	1.3	7	18.4
novelty	68	8.3	19	79.2	36	4.6	13	34.2
familiarity	48	5.9	18	75	54	6.9	20	52.6
Criteria related to the participant's beliefs and preferences	**111**	**13.6**			**128**	**16.3**		
affectiveness	98	12	22	91.7	88	11.2	27	71.1
serendipity/ curiosity	7	0.9	5	20.8	21	2.7	9	23.7
habit	4	0.5	4	16.7	14	1.8	6	15.8
time constraint	2	0.2	2	8.3	5	0.6	5	13.2
Criteria related to the quality aspects of the video or the source providing the video	**124**	**15.2**			**67**	**8.5**		
quality of source	36	4.4	15	62.5	36	4.6	13	34.2
content quality	44	5.4	18	75	17	2.2	10	26.3
technical quality	19	2.3	6	25	14	1.8	7	18.4
layout	25	3.1	14	58.3	0	0	0	0
Criteria related to audio/visual features of the video	**54**	**6.6**			**37**	**4.7**		
cinematography	17	2.1	10	41.7	6	0.8	4	10.5
visual appeal	21	2.6	14	58.3	11	1.4	5	13.2
sound/voice	16	2	10	41.7	20	2.5	10	26.3
Criteria related to the accessibility of the video	**23**	**2.8**			**11**	**1.4**		

cost	0	0	0	0	1	0.1	1	2.6
language/subtitle	10	1.2	6	25	7	0.9	4	10.5
version	13	1.6	5	20.8	3	0.4	3	7.9
Criteria related to other information within the environment	**20**	**2.4**			**17**	**2.2**		
availability	0	0	0	0	2	0.3	2	5.3
verification	5	0.6	4	16.7	8	1	5	13.2
unusualness	15	1.8	10	41.7	7	0.9	5	13.2
Criteria related to other people's opinions or you Tube's recommendations	**68**	**8.3**			**111**	**14.1**		
rank order	8	1	8	33.3	17	2.2	9	23.7
popularity	20	2.4	9	37.5	22	2.8	9	23.7
recommended video	40	4.9	15	62.5	72	9.1	20	52.6

Table 6.7 Comparisons of the mentions of relevance criteria between diaries and interviews studies

In order to determine statistical differences between criteria mentions in the two studies, Chi-squared test was conducted. The null hypothesis to be tested is "there are no significant differences in applying relevance criteria between the diary study and the interview study". Criteria with low-frequency counts (less than ten) were excluded (Time constraint, Cost and Availability). In addition, Layout criterion was also excluded as it was not in the coding scheme of the diary study. Taken as an overall, the test showed a significant difference in the number of mentions of relevance criteria between the two studies, $\chi 2\ (24) = 92.62\ P < .001$. A follow-up post hoc test was performed following the 'calculating residuals' approach to identify the criteria which contribute to the significant variance between the two studies (Sharpe, 2015). As mentioned in the previous section, the Bonferroni correction is suggested when the number of comparisons is high (comparing diary and interview studies for 25 relevance criteria) to avoid type I error (Sharpe, 2015; Macdonald & Gardner, 2000).

A Bonferroni correction was conducted and the corrected alpha was $\alpha = .002$ Table 6.8 shows the criteria that significantly varied between the two studies.

Criteria	χ^2
background experience	13.10
content quality	12.04
recommended video	10.37
novelty	10.05

Table 6.8 Relevance criteria significantly varied between the two studies

The criteria mentioned in Table 6.8 (except Recommended videos) have more mentions in the interview study compared to the diary. While applying correction to the alpha is recommended for multiple comparisons to avoid type I error, it increases the probability of missing some significant differences (type II errors). For the comparison made in this section between diary and interviews in terms of the mentions of relevance criteria, more criteria would be significantly different between the two studies if Bonferroni correction was not applied. The criteria are: Topicality, Cinematography, Recency, Habit and Version. Having collected the data from two different studies and from different samples, I do not expect an exact match between the frequencies of relevance criteria mentions between the two studies. However, the main goal of this comparison is to figure out whether the main trends in using relevance criteria for judging videos in leisure context remain steady between the two studies. Despite the difference in the frequencies of mentions of some criteria between the two studies, the ranking of the top criteria remains steady between the two studies. Topicality is the most mentioned criterion followed by Affectivness in both studies. People in the video and Familiarity both have very similar percentages of mentions and ranked in the fourth and fifth positions of ranking all the criteria in both diaries and interviews. Thus, core findings are similar in both studies. These results indicate that applying different methods (naturalistic diary study and recorded session with interviews) revealed similar and consistent findings. This adds more robustness to the findings and serves as a form of method and data triangulation (Creswell & Miller, 2000).

6.5.4 Relevance Criteria Changes between the Start and End of the Sessions

The previous sections investigated relevance criteria for each video viewed in the search session at two stages: before viewing the video (selection stage) and while viewing the video (viewing stage). As participants watch several videos during their leisure search session, they go through several episodes of selection-viewing stages. Section 6.5.2 presented the changes in criteria selections between these two stages.

In this section, I will further investigate whether relevance criteria change between the beginning and the end of the search sessions. The goal here is to examine whether the mentions of relevance criteria used at the end of the search sessions are similarly distributed to those mentions of relevance criteria at the beginning of the search sessions. This will indicate whether the participants (as a group) were consistent in applying their relevance judgment criteria through the session. Understanding how participants behave in the search session is important for IR developers as session is a key element which developers focus on (He, Göker, & Harper, 2002; Neelima & Rodda, 2016). To achieve this goal, criteria mentioned at the first and last videos (for each of the selection and viewing stages) for each session are extracted from the interviews. As I just taking the first and last videos, the sample size became small for both stages as illustrated in Table 6.9 and Table 6.10. The null hypothesis to be tested is: "there are no differences in applying relevance criteria between the beginning and end of the search sessions".

Criteria	**Start of the session**	**End of the session**
affectivness	2	2
background experience or personal memories	2	3
coverage	5	1
familiarity	14	4
genre	2	0
habit	1	0
length	5	2
novelty	7	2

people in the video	7	3
popularity	5	1
content quality	1	1
quality of source	5	1
rank order	1	2
recency	2	1
recommendation	7	7
serendipity/curiosity	2	1
sound	1	0
topicality	14	10
version	0	2
visual appeal	4	4
total	87	47

Table 6.9 Selection stage: comparisons of the mentions of relevance criteria between the start and end of the sessions

Criteria	**Start of the session**	**End of the session**
affectivness	19	8
background experience or personal memories	2	5
cinematography	3	0
coverage	6	3
familiarity	1	0
genre	1	0
habit	1	1
language	2	1
layout/presentation	4	8
length	1	2
novelty	5	6
people in the video	4	2
content quality	10	4
quality of source	2	0
technical quality	3	1
recency	3	1
sound	5	1
topicality	7	8
unusualness	3	2
verification	1	0
total	83	53

Table 6.10 Viewing stage: comparisons of the mentions of relevance criteria between the start and end of the sessions

In the case of small sample size, the Chi-squared test assumption of having less than 20% of cells with expected values less than five have been violated and the reliability of the test is doubted in such cases (Little, 2013; McHugh, 2013). Fisher's exact is an

alternative of the Chi-squared test and should be applied when the sample size is small or the data is sparse or unbalanced as it does not require a large sample and is based on calculating exact P values (Field, 2013; Little, 2013; Mehta & Patel, 2012). Fisher's exact test was conducted to investigate the stability of relevance criteria between the beginning and the end of the sessions for selection and viewing criteria data sets. The test showed no significant differences between applying relevance criteria at these two points of the sessions for both selection (P= 0.669, Fisher's Exact test) and viewing (P= 0.469, Fisher's Exact test) data sets. This result indicated that participants as a group were consistent in applying relevance criteria during the search session and that there were no preferences for particular criteria between the beginning and end of the session. This result implies that retrieval systems in leisure contexts are not required to support the user differently between the beginning and end of the session.

6.6 Limitations

This study also has its limitations. A limitation of the methodology can be found in the time constraint of the search sessions, some participants tended to avoid selecting long videos as they want to provide more videos in their sessions. To mitigate the effect of time constraint and prevent its influence on the study results, when participants mention the length of the video as a reason for selection, they were asked whether the study settings was the reason or whether they will have the same decision if they were not doing the search for a study experiment. Mentions of length because of the study's time limit were not counted as mentions of Length criterion.

Furthermore, in this study, all the participants' searches were conducted by a single desktop computer. YouTube's recommendation algorithm is not as good as when a person is using his personal computer. The place of conducting the search session is not the normal place where the participants usually search. To mitigate the effect of the study location, the participants were left alone to search in a private room and at

the end of the interviews, they were asked whether they experienced any difficulties because of the place and whether they have searched similar to what they would do normally. None of the participants mentioned inconvenience because of the place. Moreover, the pre-search chatting attempted to help in putting the participants in the study context by letting them describe what they usually search for on YouTube for leisure purposes. The study attempted to examine whether relevance criteria change between the beginning and the end of the search sessions. Further research might investigate other factors that might affect the participants' selection of relevance criteria such as the topic of the search session. Finally, difficulties emerged in applying quantitative approaches to triangulate the findings of two separate qualitative analyses. The conservative alpha correction applied is recommended in the case of multiple comparisons, however, it increases the probability of missing some significant differences (type II errors).

6.7 Chapter Summary

In this study, 24 participants were asked to search YouTube for leisure purposes, the sessions were screen recorded. Following the search sessions, semi-structured interviews were conducted and participants watched back their sessions and were asked about the relevance criteria they applied in judging the retrieved videos. The main contribution of the study was to investigate how users' selections of relevance criteria change through progressing in the search process for video/leisure contexts searches. This was done to answer RQ3: What is the difference in employing relevance criteria between the selecting and viewing stages of video/leisure contexts search? Investigating the dynamic aspects of relevance criteria in leisure/video contexts inform the design of video retrieval systems. Previous works of dynamic use of relevance criteria were mainly focused on academic and work-related context and mainly for text retrieval. I found that criteria selections changed at different stages of the search process. Criteria such as Recommended Video and Familiarity are crucial in the selection stage while others e.g. Affectiveness and Content Quality are more important at the viewing stage.

A secondary aim of the study was to examine the effect of the data collection methods on the findings. This was done to answer RQ4: Do different research methods provide different or similar findings? Does the diary method provide different findings from recorded search sessions with interviews?

The relevance criteria mentions resulted from a previous diary study were compared to the equivalent mentions in this study. The comparisons showed consistent results between the two studies which strengthen the findings of the list of relevance criteria for videos in leisure contexts. The following chapter will summarize and discuss the main findings reported in both Chapter 5 and 6.

Chapter 7

Discussion

This chapter discusses the research findings presented previously in Chapter 5 and 6. Chapter 5 reported on the findings of the first main study in this research. The study followed a naturalistic approach in which diary was applied as the data collection method. The goal of this study is to investigate criteria users apply in making relevance judgment decisions when searching videos in a leisure context and to what extent do these criteria match the criteria mentioned in the previous literature of text retrieval and/or work task contexts? Investigating relevance criteria for videos in leisure context study's findings paved the way for a second study where more controlled approach is followed: recorded search session followed by a semi-structured interview. Chapter 6 report on the finding of the second study. The goal of the second study is to investigate the dynamic evolution of these criteria between different search stages. Specifically, examine the differences in applying relevance criteria at two different stages of the search process (selecting and viewing). Given that the studies followed different approaches (naturalistic in the investigating relevance criteria study and controlled in the dynamic use of relevance criteria study), the second study also aims to compare the findings revealed from both studies to examine the effect of the research method used on the findings.

This chapter summarises the findings of the two main studies and discusses the implications of the findings on the design of video retrieval systems used for leisure purposes. The chapter is structured as follows: Section 7.1 discusses the topics and motivations of leisure searches as arises from this research, mainly from investigating

relevance criteria study in Chapter 5. Section 7.2 reflects on the relevance criteria for searching for videos in leisure contexts reported in Chapter 5. The findings of the dynamic use of relevance criteria study (Chapter 6) are discussed in Section 7.3. In Section 7.4, the design implications of the findings are presented.

7.1 Topics and Motivations for Video Leisure Searches

Both studies showed that participants search for a variety of topics and triggered by various motivations when searching YouTube for leisure purposes. Section 5.5.3 in Chapter 5 classified the topics participants searched for into labels as demonstrated in Table 5.3. The same search topics were reported by the participants in the dynamic use of relevance criteria study. Comparisons between the topics mentioned in this research and those mentioned in previous video retrieval studies for leisure purposes (Cunningham & Nichols 2008; Yeh 2016) showed similarity.

The majority of topics mentioned in both studies could be classified as casual leisure searches specifically the passive entertainment category of casual leisure. Examples of these searches include listening to a song or watching a movie trailer. On the other hand, few searches in the data provided in this research could be related to serious leisure where the participants are keen to know more about a specific serious leisure activity they usually do. For example, one of the participants in the dynamic use of relevance criteria study searched for surfing videos as it is a particular hobby he likes to practice in his free time.

> "I go surfing maybe every three weeks. So when I don't go, I like to watch it. Because I'm kind of sick of the city sometimes. I like nature. And during the weekend I don't really have time to go outside too much, so I think this is…" P4

Another participant has the hobby of possessing and playing professional drum machines. So he spent the search session looking for reviews on synthesisers and

drum machines. “The reason that I was interested in it was because I have this synthesiser, this Arturia Microbrute” P24.

Yang's study of video relevance criteria for work-related tasks found that the participants' video information needs are complicated because users might need to search videos based on visual information, audio information or combination of both as provided by the retrieval system. An example of combination need, “basket players on the court with audience cheering” (Yang, 2005). The participant might also be interested in a specific scene in the video. In this research, participants’ needs could be for audio (song or music) or visual (Europa Park) information and the need is considered as high level. Yang's participants reported specific needs (a video of shaking hand). In this research, the information need is less identified and about a general topic e.g. funny videos, healthy food recipe. Thus, in leisure context, the participants are less able to precisely define the information needs and what would be relevant to their needs unless the cases where participants are relooking for familiar videos they have watched before or recommended videos by friends e.g. "Meghan Trainor’s new single video celeb ‘NO’". This is in alignment with casual leisure information behaviour model (Elsweiler et al., 2011) which stated that one of the qualities that distinguished casual leisure information scenarios from work-based scenarios is the type of needs. The needs in the casual-leisure scenario are vague and there are no consequences for failing to meet these needs. In other words, there is no activity that might fail to complete if the information need is not met.

In 38% of cases reported in the diary study, information needs are totally absent. The participants did not report on information needs and the leisure search sessions were triggered by various motivations as illustrated in Table 5.4 in Section 5.5.3. The motivations mentioned in the diary study also overlap with the motivations mentioned in previous casual leisure video viewing studies (Cunningham & Nichols 2008; Yeh 2016). Examination of these motivations revealed that many of them (e.g. pass time, change mode) are not related to finding information. Hedonism is the common property for most of those motivations. Again, this is in alignment with the

casual leisure information behaviour model (Elsweiler et al., 2011) which stated that the experience is more important than finding information and that many of casual leisure motivations have little relation to information.

Finally, the information needs and motivations reported in this research correspond to some of Stebbins's benefits for pursuing casual leisure activities (Stebbins 2007). For example, one of these benefits is edutainment which indicates that participants could learn something during their casual-leisure activities. There is evidence in the data that participants watching videos not only for entertainment but also to learn new things. Examples include: "I want to learn new techniques in makeup" and "know new ways of organizing". Edutainment motivations are labelled as "learning" in the participants' motivations for leisure search in Chapter 5 Table 5.4. Regeneration or re-creation is another benefit of casual-leisure which is obvious in the participants' motivations such as "to brighten mood" and "to ease my mind and get the mood to sleep". Such motivations are grouped under "change mood" and "relaxation and refreshment" in this research. Well-being is also achieved by the types of searches participants conducted in this research.

7.2 Relevance Criteria for Searching Videos in Leisure Contexts

Investigating relevance criteria for videos in leisure context study applied a diary study in order to address RQ1. Previous studies in the area of relevance criteria focussed mainly on academic or work-related contexts and the information judged were mainly in the text format. Thus, this study attempted to fill in the gap of how people assess videos to watch for leisure purposes. The study also aims to compare the findings with this previous literature in order to address RQ2: To what extent do these criteria match the criteria mentioned in text retrieval and/or work task context?

The findings revealed 28 relevance criteria participants apply when searching videos for leisure purposes and those criteria grouped into eight categories as stated in Chapter 5 Section 5.5.1. This section summarizes the main findings of the study and discusses the findings in relation to previous literature of relevance criteria.

It has always been challenging to compare the findings of relevance criteria studies for several reasons. One reason behind this difficulty is that different studies have various labels and definitions for similar relevance criteria and the grouping/categorization of the findings also varied among different studies. Bales and Wang (2006) attempted to synthesize relevance criteria from 16 relevance studies; however, the work covered only a limited number of studies. Another challenge in comparing relevance criteria studies is that various methodologies have been applied in the relevance criteria literature (Maglaughlin & Sonnenwald, 2002; Savolainen & Kari, 2006).

Investigating relevance criteria for videos in leisure context study revealed that many relevance criteria mentioned in the academic or work-related contexts can be used in the leisure contexts too. As mentioned in Chapter 4, the coding scheme used in this research inherited many labels from the previous literature. Thus, there is an overlap between the relevance criteria mentioned in this study and the previous studies with new criteria emerged from the data analysis of the diaries.

The study showed that criteria related to the information content of the videos are the most popular category of criteria, with Topicality being the most dominant criterion that contained almost 20% of the overall mentions of relevance criteria. This is in alignment with the previous literature which also found it to be a dominant criterion. Thus, Topicality retains its position as a core and dominant criterion even when moving among contexts, in specific, moving from academic/work-related contexts to leisure contexts I also found Topicality to be an essential criterion. And I was able to expand the scope of Topicality.

The definition of Topicality in this study exceeds (Chapter 5 Section 5.5) the limited definition of the match between the participant's query and the retrieved videos to include responses where the video match the participant's interest as well. For example, "the topic is interesting". Some other research (Hirsh, 1999; Reuter, 2007) have a standalone code for Interest. Hirsh defined Interesting criterion as "Texts/Pictures capture the student's attention and generate curiosity" while Reuter defined it as "mention of general interest in the book". Hirsh's definition might cause confusion as it mixed up between Interest and Curiosity and Reuter based her definition on simply mentioning Interest in the participant's comment such as "I want to see what that one's about. It looks *interesting*". The analysis I followed in this research depends on understanding the underlying meaning of the participants' utterances rather than simple direct mention of the relevance criteria in the utterances. Moreover, as also Hirsh noticed, the mention of Interest is basically indicated how personally the participant thinks about the retrieved object's topic.

> "The liberal use of this relevance category for text and pictures indicates the extent to which children personalized the research topic. By mentioning that the textual information or graphical image was interesting, they were really saying that it was interesting to them personally" (Hirsh, 1999)

So Interest as a criterion is tightly related to the video topic. Thus, to avoid confusion with other relevance criteria such as (Curiosity and Novelty) and because of the relation between Interest and the topic of the video, I augmented the definition of Topicality in this research to be "the extent to which information provided in the video matches the participant's search topic or interest".

Within the criteria related to the information content category, People in the video is a new code emerged from the analysis of the diaries. Previous studies (Barry, 1994; Yang, 2005) mentioned the author or the participant's relationship with the author as a criterion in making relevance judgement decisions. As video retrieval is different than text and it is hard to identify the author of a video, I found the Author code not

suitable in the study case and a more general definition is needed which include not only the author but also persons appeared in the video (TV host, singer, actor, band, YouTuber or guest, etc.).

The second most mentioned category of relevance criteria were those related to the participants' beliefs and preferences or situations. Although include some different criteria, this group was also ranked third in Barry's study (Barry, 1994). Affectivness is the criterion in this group that acquired the second top mentions after Topicality. The increase of importance of Affectivness is mainly because of the change in the studies' contexts. Moving from academic or work-related context to leisure context has affected the importance of this criterion that used to play a less important role in the participant's relevance judgement decision in the former context.

Within this category of relevance criteria, Habit was a new code that can be considered as a sub-label to Familiarity but with the restriction of mentioning watching a video in a repetitive manner or as part of another habit. This code could not be applicable in previous academic or work-related relevance criteria studies. Habit was mentioned as a motivation for casual leisure search in (Elsweiler et al., 2011) diary study. In this research, I found that Habit is applied in relevance judgment as well.

Criteria related to other people's opinions or YouTube recommendations was the third top mentioned category of criteria, including Recommended video which is a partially new code. Some previous studies such as Westman and Oittinen (2006) have a Recommendation from Others criterion in their study; however, Recommended video in this study has a more broad meaning. In addition to recommendation provided by friends, the participants also mentioned recommendations by YouTube, web pages, advertisements in public places and other social media sites (Twitter, Instagram). Moreover, Recommendation criterion was a marginal criterion in Westman and Oittinen (2006) study with few mentions, while it is one of the top

mentioned criterion in this research. This indicates the importance of recommendations in leisure context compared to the work-related search context.

There were a group of relevance criteria from previous studies in academic context which did not appear in my data. Examples of these criteria include: Affiliation "relating a document to a particular organization, event, or initiative" (Fulton & Vondracek, 2009), Authority which is a judgement based on references or citations and Literature "whether the video has related literature" (Yang, 2005). These criteria tend to be more specific to the academic context and could not be applied in other contexts such as leisure.

Moving from text retrieval to video retrieval, criteria related to audio/visual features of the video emerged such as Cinematography, Visual Appeal, Sound/Voice in addition to Technical Quality.

To conclude, there is a considerable overlap between leisure relevance criteria and previous relevance criteria studies with the importance of the criteria (indicated by the number of mentions) varying among different contexts. New criteria emerged which tend to be more related to the leisure context.

7.3 Dynamic Aspects of Relevance Criteria

The main contribution of the dynamic use of relevance criteria study was to examine how users' selections of relevance criteria change through progressing in the search process for video/leisure contexts searches. This is to address RQ3: What is the difference in employing relevance criteria between the selection and viewing stages of video/leisure contexts searches? This investigation was conducted with the goal of providing additional insights into the dynamic use of relevance criteria. A secondary aim of the study was to examine the effects of changing the methods of data collection on the findings. The two studies that form this research followed two different approaches, the first study collected data in a naturalistic setting using

diaries while the second study applied more controlled setting using recorded search sessions followed by interviews. Comparing relevance criteria resulted from both studies in order to examine methods effects should address RQ4: Does the diary method provide different findings from recorded search sessions with interviews? The remainder of this section will discuss the dynamic use of relevance criteria study's findings in relation to previous dynamic relevance criteria studies.

The findings reported in Chapter 6 showed that Topicality, Familiarity, Recommended video, People in the video and Novelty are the top five criteria in the selection stage. In the viewing stage, the top five mentioned criteria changed to Affectivness, Topicality, Content quality, Novelty and Layout. Moreover, there is a significant difference in applying relevance criteria between selection and viewing stages. This is in alignment with some of the previous dynamic relevance criteria studies e.g. (Reuter, 2007; Tang & Solomon, 2001; Wang & White, 1999; Xie & Benoit, 2013)

The study found that criteria such as Topicality and Familiarity play an important role in the participant's initial relevance judgment at the selection stage. Participants are first attracted to videos which are familiar to them or from familiar channels. In this early stage, participants are also attracted to videos with appealing thumbnails and popular videos. Mikkonen and Vakkari, (2016) have also found familiarity the most mentioned interest criteria for selecting novels.

As the search progresses, the importance of Familiarity decreased for the sake of Affectiveness, which became the most dominant criterion at the viewing stage. This is in alignment with a previous study about recreational reading, which found Familiarity less prevalent at later stages of the search (Reuter, 2007). Participants became more specific and apply other criteria regarding the content of the video and how the information was presented at this stage. In addition, audio/visual criteria such as Cinematography, Sound and Technical Quality increased significantly.

These findings indicate that although participants might initially base their judgments on the topic of the video or their familiarity with the video or the channel, as they progress in the search other criteria such as Affectiveness and Quality content became more crucial. Affectiveness was the criterion with the highest significant change between the two stages. This in lie with Mikkonen and Vakkari, (2016), who found that user's previous knowledge and novels' title are the initial triggers to users' interest in a novel, content description contribute to the final decision for selecting a novel.

In line with Xu (2007), Novelty found to be an important criterion for non-problem solving task at the selection and viewing stages of the search. Novelty remains steady while moving between selection and viewing stages in this study. Previous studies did not agree about the development of the importance of Novelty among search stages. Tang and Solomon (2001) study found Novelty to be more important at later stages while Reuter (2007) reported an increase in mentions of Novelty at earlier stages of the search.

Based on previous research, it was expected that Topicality would experience a decrease in the number of mentions moving from selection to viewing stages. These study results confirm the findings from previous studies, Topicality became less important as the participants move to the viewing stage.

Besides, examining the change in relevance criteria between selection and viewing stages for each video, the study examined the use of relevance criteria between the beginning and end of the search sessions, the findings showed stability in applying relevance criteria between the start and end of the sessions. Some of the previous work (Hirsh, 1999; Taylor, 2009) in academic related context found changes in applying relevance criteria as search session progress and relate this to change in user's cognitive state. Users, as they progress in their searches, gain more understanding of the search task and topic. This is not always the case in leisure search. This result indicates that leisure search context is different, participants do

not necessarily start their searches with vague cognitive state and unfamiliarity of the search tasks. As in leisure context, the goal is not always to fill in knowledge gap, the main aim of the search is entertainment. Even when the participants search for specific topics (e.g. airlines crafts, surfing videos) they are still keen to apply a finite set of criteria as they progress in their search sessions.

The relevance criteria resulted from the analysis of the interviews in the dynamic use of relevance criteria study were compared to those resulting from the analysis of the diaries in the investigating relevance criteria for videos in leisure context study. In general, the findings (in terms of relevance criteria identified) of both studies are consistent with each other where both studies reported on the same dominant criteria (e.g. Topicality, Affectiveness). Only two criteria found significantly different between the two studies. Few studies in the literature have compared diary study results to other research methods, thus, this study contributes to comparison of methods.

7.4 Implications for System Design

Video retrieval systems such as YouTube provide some search filters that support the relevance criteria that users employ to judge the relevance of the videos. For example, the data showed participants have preferences based on the Length of the video (long or short videos), Popularity, Recency and the Technical quality (HD versions). YouTube offers filters to search for videos that are less/greater than 20 mints, rank the result list by date of upload and also facilitate the search for videos in a specific format (HD for example). Similar to previous studies which found that advanced search options are rarely used (Choi, 2010) and their use is not intuitive (Taylor, 2009), none of the participants in the dynamic use of relevance criteria study apply any of these filters even when they met their relevance criteria. It might be that participants are not aware of their existence or they were hidden. The study confirms Choi's findings that such filters should be made easily reached on the main search page to encourage the users to get benefits from them. So maybe it is worth that

video search engines identify the relevance criteria that are crucial to the user based on his search history and inform him with the advanced search tools suitable to him. It might be a small ad at the beginning of the video or a pop-up message that shows in brief what search filter would enhance search result and suitable to the searchers' criteria.

As the findings showed that in leisure searches participants mostly showed consistency in applying relevance criteria between the beginning and end of the session. This finding indicates that video retrieval developers might treat session as the basic unit of analysis rather than user profile in general.

Furthermore, the findings showed low mentions of Serendipity/Curiosity as a relevance criterion. It is an open question to explore whether serendipity is not required by users in leisure searches or the video retrieval system are not supporting the users enough to serendipitous encountering of interesting videos. This study did not answer this question, however, it suggests the investigation of the role of serendipity in video leisure search and how could it be improved. Although the role of serendipity in casual leisure search is not the main goal of this research, looking at serendipity concept in both casual leisure and recommender systems studies would provide some reflections on this finding. By looking at recent studies of recommender systems, it is evident that current recommender systems are achieving satisfying levels of quality in recommendations provided to the user based on accuracy (de Gemmis, Lops, Semeraro, & Musto, 2015). In other words, the systems are able to suggest accurate items based on the user's search profile or other users with similar interest. However, studies (Kotkov, Wang, & Veijalainen, 2016; Trattner & Elsweiler, 2017) have shown that these systems lack the support of serendipity (enabling the user to surprisingly finding interesting information) and emphasize the importance of taking serendipity into account as a factor for evaluating the quality of the recommendations. On the other hand, casual leisure is also known as the main source of serendipity (Stebbins, 1997). Thus, having low mentions of serendipity in this research might indicate that serendipity is required but should be better supported

in the recommendations provided by YouTube. While this research did not attempt to evaluate the recommendation algorithms applied by YouTube, only explicit users' mentions of serendipity are considered. Participants were not required to evaluate each recommendation based on serendipity.

7.5 Chapter Summary

This chapter has discussed the findings of the two main studies that form this research. It presents interpretations of the findings and discusses them in light of previous literature. The chapter also provides the implications of these findings in terms of system design recommendations. The following chapter concludes the thesis by summarizing the main findings and describing how the research questions are met. The chapter will also present suggestions for future work.

Chapter 8

Conclusion

This chapter presents an overview of this research. First, the chapter restates the research problem and how the research questions addressed. The main findings of the research will be also presented. Then, the chapter highlights the contributions to knowledge this research makes and provide suggestions for future work.

8.1 Addressing the Research Problem and Questions

The focus of this research is on the user relevance area. The main objective is to examine how people judge the relevance of videos watched for leisure purposes. The research aims to investigate relevance criteria people apply in their relevance judgment decision and how these criteria might change at different stages of the search process.

Previous studies in user relevance criteria were mainly focused on text retrieval in academic or work-related contexts. Little is known about how people judge videos watched for leisure purposes. As more content online is in the form of images and videos it worth to investigate relevance criteria for different media types. Furthermore, people have limited time for leisure and it is essential for them to make better use of this time. Many searches nowadays are for leisure purposes (Elsweiler, Wilson, & Harvey, 2012). Thus, a research which investigates the use of relevance criteria when searching for videos in leisure contexts is needed to figure out the relevance criteria used in this different context and media.

Two studies were conducted to achieve the goal of the research. In the first study, a naturalistic approach was followed by using a diary study. Thirty participants completed diaries for the duration of one week. In the diaries, the participants were reporting on the relevance criteria they applied when selecting and watching videos on YouTube for leisure purposes.

The second study followed a more controlled approach by recording participants' leisure search sessions and conducting interviews afterwards to discuss relevance criteria they applied during relevance judgment decisions. The study examined the differences in applying relevance criteria at two different stages of the search: selection and viewing.

The thesis aims to address the following research questions:

- RQ1: What are the relevance criteria users apply when judging videos in a leisure context?
    - Subsequently, which relevance criteria are the most important when judging videos in leisure context?
- RQ2: To what extent do these criteria match the criteria mentioned in the previous literature of text retrieval and/or work task context?
- RQ3: What is the difference in employing relevance criteria between the selecting and viewing stages of video/leisure contexts search?
    - Subsequently, are there significant differences in applying relevance criteria between the selecting and viewing stages of video/leisure contexts search?

- RQ4: Do different research methods provide different or similar findings? Does the diary method provide different findings from recorded search sessions with interviews?

These research questions were motivated by the knowledge gap found in the literature. Specifically, the fact that previous relevance criteria studies have concentrated on academic or work-related context and the objects retrieved were mainly in text format. Thus RQ1 aims to figure out what is the state of relevance criteria beyond academic/work-related contexts and beyond text retrieval. RQ2 follows on RQ1, it aims to compare the findings with previous literature to examine the similarities and differences between the two contexts and between text and video retrieval. RQ3 was motivated by the fact that relevance is dynamic and relevance criteria selection might change among different search stages or as the search progress in time. So this research question aims to investigate the dynamic aspects of relevance criteria for video searching in leisure contexts. It investigates the differences in applying relevance criteria between two stages of the search process: selection and viewing. The study also investigates the change in applying relevance criteria between the begging and end of the search sessions. Given that this research has collected data about relevance criteria using two different approaches: naturalistic and controlled approaches, RQ4 aims to compare the findings of both studies to examine whether the findings were consistent between the two studies or different methods lead to different findings.

To address RQ1, the investigating relevance criteria for videos in leisure context study was conducted (Chapter 5) and participants' inputs regarding their relevance criteria were analyzed. The analysis revealed 28 video relevance criteria in leisure context and these criteria were grouped into eight categories. The study found that criteria related to the information content of the video were the most dominant category. Of the individual criteria, Topicality, Affectiveness and Recommended video were used more frequently in relevance judgements. New criteria emerged from the diaries analysis, for example, People in the video is a new criterion within the criteria related

to the information content category. People in the video criterion includes not only the author but also persons appeared in the video (TV host, singer, actor, band, YouTuber or guest, etc.). Another new criterion related to the participants' beliefs or preferences is Habit. Habit is a new label that can be considered as a sub-label to Familiarity but with the restriction of mentioning watching a video in a repetitive manner or as part of another habit. Criteria related to audio/visual features of the video emerged such as Cinematography, Visual appeal, Sound/Voice in addition to Technical quality.

To address RQ2, the discussion provided in Chapter 7 compared the findings of the investigating relevance criteria for videos in leisure context study to the previous literature of relevance criteria. The comparisons revealed an overlap between relevance criteria mentioned in this study with previous literature and that many criteria mentioned in the previous text or academic/ work-related contexts still applied to video and leisure contexts. However, the importance of these criteria varies when moves between the two contexts. New relevance criteria also emerged from the data and these criteria tend to be more specific to leisure contexts.

RQ3 was addressed in the dynamic use of relevance criteria study (Chapter 6). Relevance criteria were investigated in two stages of the search process: selection and viewing. The stages were adopted from Yeh's (2016) stages of casual leisure searching for videos online. Statistical comparisons between the relevance criteria applied in the selection and viewing stages were conducted. The analysis revealed criteria selections changed at different stages of the search process. Criteria such as Recommended video and Familiarity are crucial in the selection stage while others e.g. Affectiveness and Content quality are more important at the viewing stage.

RQ4 was also answered in Chapter 6 Section 6.5.3. The relevance criteria mentions resulted from the relevance criteria investigation study (Chapter 5) were compared to the equivalent mentions in the dynamic use of relevance criteria study (Chapter

6). The comparisons showed consistent results between the two studies which strengthen the findings of the list of relevance criteria for videos in leisure contexts.

8.2 Contributions

The findings from this research make the following contributions to the user relevance research area and information retrieval systems.

1. The thesis uncovered the use of relevance criteria when searching for videos in leisure context and compares and contrasts the findings with relevance criteria emerged from previous literature. The full list and definitions of relevance criteria for video/leisure searches are provided in Chapter 5 Section 5.5.1 and comparing the criteria emerged from this research with the previous literature is presented in Chapter 7 Section 7.2. Previous studies of user relevance criteria have mainly concentrated on academic or work-related contexts and mainly for text retrieval. This research attempts to enrich the current relevance criteria literature by investigating users' video relevance criteria in leisure contexts.
2. This research uncovered the dynamic aspects of relevance criteria applied for videos searched for leisure purposes. Specifically, the thesis investigates the differences in relevance criteria at the stages of selecting and viewing videos for leisure and also examines the stability in applying these criteria between the beginning and end of the search sessions. Previous works in dynamic use of relevance criteria were more on academic contexts where participants are students who need information to complete their assignments and data were mainly collected at different phases of the assignment. Leisure is a different context, thus the dynamic use of relevance criteria when searching for videos for leisure purposes was investigated. By shedding the light on this uncovered area of research and by examining the dynamic use of relevance criteria for videos, the IR community will gain a deeper understanding of how users make their

relevance judgment decisions in leisure context. Furthermore, the findings of this research have implications on the design of IR systems as shown in Section 7.2.2

3. The thesis provides a unified coding scheme based on several relevance criteria studies to produce a comprehensive coding scheme for leisure/video searches as shown in Chapter 4 Section 4.3. Previous relevance criteria studies either consider only one well-known coding scheme such as Barry and Schamber or design their own coding scheme resulting in the problem of redundant between relevance criteria studies.
4. The thesis provides methods comparisons by comparing the findings of the main two studies that form the thesis to examine whether the method used affects the study findings. The two studies followed two different approaches: naturalistic approach was followed in the first study by using diaries and a more controlled approach in the second study by recording participants' search sessions followed by interviews. Few studies in the literature have compared diary study results to other research methods, thus, this research contributes to methods comparisons.

8.3 Future Work

Further studies are required to fully investigate the use of video relevance criteria in leisure contexts. Starting from the limitations of the two studies that form this research, participants sample was one of these limitations. The majority of the participants were university or college students, further research could explore the use of relevance criteria for diverse groups of YouTube users (students, employees, unemployed etc.) and compare the criteria emerged from each group.

There were some mentions from both studies indicating that watching one video influences the behaviour and relevance criteria for the other, for example, "this video shape the nature of the coming video I am going to watch". Future works might further investigate this issue to link relevance criteria applied on a video to the next.

The second study of this research focused on the dynamic use of relevance criteria in leisure search in general, future research might investigate in more depth the differences in dynamic applying of relevance criteria among different leisure topics. The study also suggested future research to investigate the role of serendipity in video leisure search and how could it be improved. Such investigation will aid video retrieval systems used in leisure contexts to better support the users and facilitate serendipitous encountering of interesting videos.

8.4 Closing Remarks

As Saracevic stated, “relevance is here to stay” (Saracevic, 2007b). This research has contributed to the area of user relevance criteria. It is aimed to aid in better understanding of users’ relevance judgments by investigating relevance criteria users apply when searching for videos in leisure contexts. The deeper understanding of the users’ criteria applied during the relevance judgment decisions and the dynamic aspects of them are essential to meet leisure users’ needs and to improve video retrieval systems used for leisure purposes. The studies provided in this research enhance the current literature by shedding the light on a less discussed context: leisure.

References

Adobe Digital Index. (2013). *The U . S . Digital Video Benchmark*.

Al-Harbi, A. L., & Smucker, M. D. (2014). A qualitative exploration of secondary assessor relevance judging behavior. In *Proceedings of the 5th Information Interaction in Context Symposium on - IIiX '14* (pp. 195–204). New York, NY, USA: ACM Press. https://doi.org/10.1145/2637002.2637025

Alaszewski, A. (2006). *Using diaries for social research*. London: SAGE.

Albassam, S. A. A., & Ruthven, I. (2018). Users' relevance criteria for video in leisure contexts. *Journal of Documentation*, *74*(1), 62–79. https://doi.org/10.1108/JD-06-2017-0081

Alexa. (2018). The top 500 sites on the web. Retrieved June 15, 2018, from https://www.alexa.com/topsites

Athukorala, K., Hoggan, E., Lehtio, A., Ruotsalo, T., & Jacucci, G. (2013). Information-seeking behaviors of computer scientists: challenges for electronic literature. *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*, *50*(1), 1–11.

Bailey, C. A. (2018). *A guide to qualitative field research* (3rd ed.). Thousand Oaks, California: Sage Publications.

Bailey, P., Craswell, N., Soboroff, I., Vries, A. P. De, & Yilmaz, E. (2008). Relevance assessment: are judges exchangeable and does it matter. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 667--674). https://doi.org/10.1145/1390334.1390447

Balatsoukas, P., O'Brien, A., & Morris, A. (2010). Design factors affecting relevance judgment behaviour in the context of metadata surrogates. *Journal of Information Science*, *36*(6), 780–797. https://doi.org/10.1177/0165551510386174

Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, *63*(9), 1728–1746. https://doi.org/10.1002/asi.22707

Bales, S., & Wang, P. (2006). Consolidating user relevance criteria: A meta-ethnography of empirical studies. In *Proceedings of the American Society for Information Science and Technology* (Vol. 42). https://doi.org/10.1002/meet.14504201277

Barry, C. L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science*, *45*(3), 149–159.

https://doi.org/10.1002/(SICI)1097-4571(199404)45:3<149::AID-ASI5>3.0.CO;2-J

Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, *34*(2/3), 219–236. https://doi.org/10.1016/S0306-4573(97)00078-2

Bateman, J. (1997). Changes in Relevance Criteria: A Longitudinal Study. *Proceedings of the ASIS Annual Meeting*, *35*, 23–32.

Bell, D. J., & Ruthven, I. (2004). Searcher's assessments of task complexity for web searching. In S. McDonald & J. Tait (Eds.), *Advances in Information Retrieval. ECIR 2004. Lecture Notes in Computer Science* (pp. 57–71). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-24752-4_5

Berk, R. (2009). Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. *International Journal of Technology in Teaching and Learning*, *5*(1), 1–21.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Borlund, P. (2003). The concept of relevance in information retrieval. *Journal of the American Society for Information Science and Technology*, *54*(10), 913–925. https://doi.org/http://dx.doi.org/10.1002/asi.10286

Bruce, H. W. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science*, *45*(3), 142. https://doi.org/10.1002/(SICI)1097-4571(199404)45:3<142::AID-ASI4>3.0.CO;2-6

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford, United Kingdom: Oxford University Press.

Brzozowski, M., Crossley, T. F., & Winter, J. K. (2017). A comparison of recall and diary food expenditure data. *Food Policy*, *72*, 53–61. https://doi.org/10.1016/J.FOODPOL.2017.08.012

Burgess, J., & Green, J. (2009). *YouTube: Online Video and Participatory Culture*. Cambridge, MA: Polity.

Carey, R. F. (2014). What Pictures Tell: The Future of Social Is Visual | Social Media Today. Retrieved August 28, 2015, from http://www.socialmediatoday.com/content/what-pictures-tell-future-social-visual

Cho, J. Y., & Lee, E.-H. (2014). Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. *The Qulatitative Report*, *19*(32), 1–20.

Choi, Y. (2010). Effects of contextual factors on image searching on the Web. *Journal of the American Society for Information Science and Technology*, *61*(10), 2011–2028. https://doi.org/10.1002/asi.21386

Choi, Y., & Rasmussen, E. M. (2002). Users' relevance criteria in image retrieval in American history. *Information Processing & Management*, *38*(5), 695–726. https://doi.org/10.1016/S0306-4573(01)00059-0

Chung, E. K., & Yoon, J. W. (2012). Analysis of multimedia needs and searching features an exploratory study. In *Proceedings of the ASIST Annual Meeting* (Vol. 49). https://doi.org/10.1002/meet.14504901297

Cisco Public. (2017). *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021(White paper)*.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data : complementary research strategies*. Thousand Oaks: Sage Publications.

Cole, C., Beheshti, J., & Abuhimed, D. (2017). A relevance model for middle school students seeking information for an inquiry-based class history project. *Information Processing & Management*, *53*(2), 530–546. https://doi.org/10.1016/J.IPM.2016.10.002

Cool, C., Belkin, N. J., Frieder, O., & Kantor, P. B. (1993). Characteristics of text affecting relevance judgments. In *Proceedings of the 14th National Online Meeting* (pp. 77–84). Medford, NJ: Learned Information. https://doi.org/10.1.1.93.9372

Corti, L. (1993). Social research update 2: using diaries in social research. Retrieved from http://sru.soc.surrey.ac.uk/SRU2.html

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing and Management*, *36*(4), 533–550. https://doi.org/10.1016/S0306-4573(99)00072-2

Creswell, J. W., & Miller, D. L. (2000). Determining Validity in Qualitative Inquiry. *Theory Into Practice*, *39*(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2

Crystal, A., & Greenberg, J. (2006). Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology*, *57*(10), 1368–1382. https://doi.org/10.1002/asi.20436

Cunningham, S. J., & Nichols, D. M. (2008). How people find videos. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 201–210). New York, NY: ACM Press. https://doi.org/10.1145/1378889.1378924

de Gemmis, M., Lops, P., Semeraro, G., & Musto, C. (2015). An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, *51*(5), 695–717. https://doi.org/10.1016/J.IPM.2015.06.008

Duffy, P. (2008). Engaging the YouTube Google-Eyed Generation: Strategies for Using Web 2.0 in Teaching and Learning. *Electronic Journal of E-Learning*, *6*(2), 119–130

Ellis, D. (1993). Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach. *The Library Quarterly*, *63*(4), 469–486. https://doi.org/10.1086/602622

Ellis, D., & Haugan, M. (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, *53*(4), 384–403. https://doi.org/10.1108/EUM0000000007204

Elsweiler, D., & Harvey, M. (2015). Engaging and maintaining a sense of being informed: Understanding the tasks motivating twitter search. *Journal of the Association for Information Science and Technology*, *66*(2), 264–281. https://doi.org/10.1002/asi.23182

Elsweiler, D., Ruthven, I., & Jones, C. (2007). Towards Memory Supporting Personal Information Managment Tools. *Journal Of The American Society For Information Science and Technology*, *58*(7), 924–946. https://doi.org/10.1002/asi

Elsweiler, D., Wilson, M. L., & Harvey, M. (2012). Searching 4 FUN Workshop. In *the 34rd European Conference on Information Retrieval* (pp. 1–29). Barcelona, Spain.

Elsweiler, D., Wilson, M. L., & Lunn, B. K. (2011). Understanding Casual-Leisure Information Behaviour. In J. Spink, A. & Heinström (Ed.), *New Directions in Information Behaviour (Library and Information Science* (pp. 211–241). Emerald Group Publishing. https://doi.org/10.1108/S1876-0562(2011)002011a012

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). Sage Publications.

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A Description of Think Aloud Method and Protocol Analysis. *Qualitative Health Research*, *3*(4), 430–441.

Fulton, C., & Vondracek, R. (2009). Introduction: Pleasurable Pursuits: Leisure and LIS Research. *Library Trends*, *57*(4), 611–617. https://doi.org/10.1353/lib.0.0061

Götze, E., Prange, C., & Uhrovska, I. (2009). Children's impact on innovation decision making. *European Journal of Marketing*, *43*(1/2), 264–295. https://doi.org/10.1108/03090560910923328

Graneheim, U. ., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, *24*(2), 105–112. https://doi.org/10.1016/j.nedt.2003.10.001

Gray, D. E. (2014). *Doing Research in the Real World* (3rd ed.). London: Sage Publications.

Greenberg, B. S., Eastin, M. S., Skalski, P., Cooper, L., Levy, M., & Lachlan, K. (2005). Comparing survey and diary measures of internet and traditional media use.

Communication Reports, *18*(1), 1–8. https://doi.org/10.1080/08934210500084164

Gull, C. D. (1956). Seven years of work on the organization of materials in the special library. *American Documentation*, *7*(4), 320–329. https://doi.org/10.1002/asi.5090070408

Halvey, M., & Keane, M. T. (2007). Analysis of Online Video Search and Sharing. In *HT '07 Proceedings of the eighteenth conference on Hypertext and hypermedia* (pp. 217–226). New York, NY, USA: ACM.

Hamid, R. A., & Thom, J. A. (2010). Criteria That Have An Effect on Users While Making Image Relevance Judgments. In *Proceedings of the Fifteenth Australasian Document Computing Symposium* (pp. 76–83). Melbourne, Australia: School of Computer Science and IT, RMIT University.

Hanson, Gary; Haridakis, P., & Haridakis, P. (2008). YouTube Users Watching and Sharing the News: A Uses and Gratifications Approach. *Journal of Electronic Publishing*, *11*(3). https://doi.org/http://dx.doi.org/10.3998/3336451.0011.305

Hartel, J. (2003). The serious leisure frontier in library and information science: Hobby domains. *Knowledge Organization*.

Hartel, J. (2005). Serious Leisure. In K. E. Fisher, S. Erdelez, & L. McKechnie (Eds.), *Theories of Information Behavior: A Researcher's Guide* (pp. 313–317). Medford, NJ: Information Today.

Hartel, J. (2006). Information activities and resources in an episode of gourmet cooking. *Information Research*, *12*(1), Paper 281.

Hartel, J. (2013). The Serious Leisure Perspective (SLP). Retrieved August 28, 2015, from http://www.seriousleisure.net/

He, D., Göker, A., & Harper, D. J. (2002). Combining evidence for automatic Web session identification. *Information Processing & Management*, *38*(5), 727–742. https://doi.org/10.1016/S0306-4573(01)00060-7

Hirsh, S. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science*, *50*(14), 1265–1283.

Hruschka, D. J., Schwartz, D., St.John, D. C., Picone-Decaro, E., Jenkins, R. A., & Carey, J. W. (2004). Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research. *Field Methods*, *16*(3), 307–331. https://doi.org/10.1177/1525822X04266540

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huang, M., & Wang, H. (2004). The influence of document presentation order and

number of documents judged on users' judgments of relevance. *Journal of the American Society for Information Science and Technology*, *55*(11), 970–979. https://doi.org/10.1002/asi.20047

Hung, T.-Y., Zoeller, C., & Lyon, S. (2005). Relevance judgments for image retrieval in the field of journalism: a pilot study. In E. A. Fox, J. Neuhold, E., P. Premsmit, & V. Wuwongse (Eds.), *Digital Libraries: Implementing Strategies and Sharing Experiences. ICADL 2005. Lecture Notes in Computer Science* (pp. 72–80). Berlin, Heidelberg: Springer.

Ju, B., & Albertson, D. (2014). Users' criteria of video digital libraries from a public affairs context. *Proceedings of the 5th Information Interaction in Context Symposium on - IIiX '14*, 235–238. https://doi.org/10.1145/2637002.2637031

Ju, B., & Gluck, M. (2011). Calibrating information users' views on relevance: A social representations approach. *Journal of Information Science*, *37*(4), 429–438. https://doi.org/10.1177/0165551511412030

Kim, Y., Sohn, D., & Choi, S. M. (2011). Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students. *Computers in Human Behavior*, *27*(1), 365–372. https://doi.org/10.1016/J.CHB.2010.08.015

Knäusl, H. (2012). Searching wikipedia: learning the why, the how, and the role played by emotion. In *Searching 4 Fun*.

Kotkov, D., Wang, S., & Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*, *111*(C), 180–192. https://doi.org/10.1016/J.KNOSYS.2016.08.014

Kuhlthau, C. C. (1993). *Seeking meaning : a process approach to library and information services*. Norwood, NJ: Ablex Pub. Corp.

Laplante, A. (2010). Users' Relevance Criteria in Music Retrieval in Everyday Life: An Exploratory Study. In *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR, Utrecht* (pp. 601–606). Cambridge, MA: Morgan Kaufmann Publishers.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction* (2nd ed.). Cambridge, MA,United States: Elsevier Inc.

Li, N., & Kirkup, G. (2007). Gender and cultural differences in Internet use: A study of China and the UK. *Computers & Education*, *48*(2), 301–317. https://doi.org/10.1016/J.COMPEDU.2005.01.007

Li, Y., & Belkin, N. J. (2010). An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, *61*(9), 1771–1789. https://doi.org/10.1002/asi.21359

Little, T. D. (2013). *The Oxford handbook of quantitative methods. Vol. 2, Statistical*

analysis. New York: Oxford University Press.

Lux, M., Lagger, C., & Marques, O. (2011). What Makes People Watch Online Videos : An Exploratory Study. In *CHI'11 extended abstracts on Human factors in computing systems* (pp. 1–6). https://doi.org/978-1-4503-0268-5/11/05

MacLean, L. M., Meyer, M., & Estable, A. (2004). Improving Accuracy of Transcripts in Qualitative Research. *Qualitative Health Research*, *14*(1), 113–123. https://doi.org/10.1177/1049732303259804

Madden, A., Ruthven, I., & Mcmenemy, D. (2014). A classification scheme for content analyses of YouTube video comments. *Journal of Documentation*, *69*(5), 693–714. https://doi.org/10.1108/JD-06-2012-0078

Maglaughlin, K. L., & Sonnenwald, D. H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant judgments. *Journal of the American Society for Information Science and Technology*, *53*(5), 327–342. https://doi.org/10.1002/asi.10049

Markkula, M., & Sormunen, E. (1998). Searching for Photos - Journalists' Practices in Pictorial IR. In J. P. Eakins, D. J. Harper, & J. M. Jose (Eds.), *The Challange of Image Retrieval (Electronic Workshops in Computing, eWIC)*.

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Austria: Klagenfurt.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149. https://doi.org/10.11613/BM.2013.018

McQuivey, J. (2008). How Video Will Take Over The World. Retrieved April 30, 2018, from https://www.forrester.com/report/How+Video+Will+Take+Over+The+World/-/E-RES44199?isTurnHighlighting=false&highlightTerm=1.8%2520million%2520words#

Mehta, C. R., & Patel, N. R. (2012). *IBM SPSS Exact Tests*. Cambridge, MA: IBM Corp.

Mikkonen, A., & Vakkari, P. (2016). Readers' interest criteria in fiction book search in library catalogs. *Journal of Documentation*, *72*(4), 696–715. https://doi.org/10.1108/JDOC-11-2015-0142

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis : an expanded sourcebook* (2nd ed.). Sage Publications.

Millan-cifuentes, J. D., Myrhaug, H., & Macfarlane, A. (2014). Curiosity driven search : When is relevance irrelevant ? In *IIiX '14 Proceedings of the 5th Information Interaction in Context Symposium* (pp. 279–282). New York, NY, USA: ACM.

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science and Technology*, *48*(9), 810–832.

https://doi.org/10.1002/(SICI)1097-4571(199709)48:9<810::AID-ASI6>3.0.CO;2-U

Mizzaro, S. (1998). How many relevances in information retrieval? *Interacting with Computers*, *10*(3), 303–320. https://doi.org/10.1016/S0953-5438(98)00012-5

Moshfeghi, Y. (2012). *Role of Emotion in Information Retrieval*. University of Glasgow.

Moshfeghi, Y., & Jose, J. M. (2013). On cognition, emotion, and interaction aspects of search tasks with different search intentions. In *WWW '13 Proceedings of the 22nd international conference on World Wide Web* (pp. 931–941). New York, NY, USA: ACM.

Neelima, G., & Rodda, S. (2016). Predicting user behavior through sessions using the web log mining. In *International Conference on Advances in Human Machine Interaction (HMI)* (pp. 1–5). IEEE. https://doi.org/10.1109/HMI.2016.7449167

Neuendorf, K. A. (2017). *The Content Analysis Guidebook* (2nd ed.). Los Angeles: SAGE.

Palen, L., & Salzman, M. (2002). Voice-mail diary studies for naturalistic data capture under mobile conditions. In *CSCW '02 Proceedings of the 2002 ACM conference on Computer supported cooperative work* (Vol. 02pages, pp. 87–95). New York, NY, USA: ACM. https://doi.org/10.1145/587078.587092

Paolisso, M., & Hames, R. (2010). Time Diary versus Instantaneous Sampling: A Comparison of Two Behavioral Research Methods. *Anthropology Faculty Publications*.

Pian, W., Khoo, C. S., & Chang, Y.-K. (2016). The Criteria People Use in Relevance Decisions on Health Information: An Analysis of User Eye Movements When Browsing a Health Discussion Forum. *Journal of Medical Internet Research*, *18*(6:e136). https://doi.org/10.2196/jmir.5513

Pickard, A. J. (2013). *Research Methods in Information*. London: Facet.

Pöntinen, J., & Vakkari, P. (2013). Selecting Fiction in Library Catalogs: A Gaze Tracking Study. In *International Conference on Theory and Practice of Digital Libraries* (pp. 72–83). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-40501-3_8

Reuter, K. (2007). Assessing Aesthetic Relevance: Children's Book Selection in a Digital Library. *Journal of the American Society for Information Science and Technology*, *58*(12), 1745–1763. https://doi.org/10.1002/asi

Riedl, R., Hubert, M., & Kenning, P. (2010). Are There Neural Gender Differences in Online Trust? An fMRI Study on the Perceived Trustworthiness of eBay Offers. *MIS Quarterly*, *34*(2), 397. https://doi.org/10.2307/20721434

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*,

53(2), 145–161. https://doi.org/10.1002/asi.10017

Rieman, J. (1993). The Diary Study: A Workplace-Oriented Research Tool to Guide Laboratory Efforts. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*, 321–326. https://doi.org/10.1145/169059.169255

Ross, C. S. (2000). Making choices: What readers say about choosing books to read for pleasure. *The Acquisitions Librarian*, *13*(25), 5–21. https://doi.org/10.1300/J101v13n25_02

Ruthven, I. (2005). Integrating approaches to relevance. In A. Spink & C. Cole (Eds.), *New Directions in Cognitive Information Retrieval. The Information Retrieval Series* (pp. 61–80). Dordrecht: Springer. https://doi.org/10.1007/1-4020-4014-8

Ruthven, I., Baillie, M., & Elsweiler, D. (2007). The relative effects of knowledge, interest and confidence in assessing relevance. *Journal of Documentation*, *63*(4), 482–504. https://doi.org/10.1108/00220410710758986

Ruthven, I., & Kelly, D. (2012). *Interactive information seeking, behaviour and retrieval*. London: Facet publishing.

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, *58*(13), 1915–1933. https://doi.org/10.1002/asi.20682

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology*, *58*(13), 2126–2144. https://doi.org/10.1002/asi.20681

Saracevic, T. (2008). Effects of Inconsistent Relevance Judgments on Information Retrieval Test Results: A Historical Perspective. *Library Trends*, *56*(4), 763–783. https://doi.org/10.1353/lib.0.0000

Saracevic, T. (2016). *The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really? Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 8). Morgan & Claypool Publishers. https://doi.org/10.2200/S00723ED1V01Y201607ICR050

Savolainen, R. (2009). Interpreting informational cues: An explorative study on information use among prospective homebuyers. *Journal of the American Society for Information Science and Technology*, *60*(11), 2244–2254. https://doi.org/10.1002/asi.21167

Savolainen, R. (2010). Source preference criteria in the context of everyday projects. *Journal of Documentation*, *66*(2), 70–92.

Savolainen, R., & Kari, J. (2006). User-defined relevance criteria in web searching. *Journal of Documentation*, *62*(6), 685–707. https://doi.org/10.1108/00220410610714921

Schamber, L. (1991). Users' Criteria for Evaluation in a Multimedia Environment. In *Proceedings of the 54 Annual Meeting of lthe American Society for Information Science* (Vol. 28, pp. 126–133). Medford, N J: Learned Information.

Schamber, L. (1994). Relevance and information behavior. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology (ARIST)* (pp. 33–48). Medford, NJ: Learned Information.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition∗. *Information Processing & Management*, *26*(6), 755–776. https://doi.org/10.1016/0306-4573(90)90050-C

Sedghi, S., Sanderson, M., & Clough, P. (2012). How do health care professionals select medical images they need? *Aslib Proceedings*, *64*(4), 437–456. https://doi.org/10.1108/00012531211244815

Silverman, D. (2015). *Interpreting qualitative data* (5th ed.). London: SAGE.

Silverman, D. (2017). *Doing qualitative research* (5th ed.). London: SAGE.

Smith, C. (2018a). 17 Interesting flickr Stats and Facts (February 2018) | By the Numbers. Retrieved April 29, 2018, from https://expandedramblings.com/index.php/flickr-stats/

Smith, C. (2018b). 17 Interesting flickr stats and facts (May 2018) by the numbers. Retrieved May 1, 2018, from https://expandedramblings.com/index.php/flickr-stats/

Smith, D. H., Stebbins, R. A., & Michael, D. (2010). A Dictionary of Nonprofit Terms and Concepts. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *21*(3), 462–463. https://doi.org/10.1007/s11266-009-9118-8

Smithson, S. (1994). Information retrieval evaluation in practice: A case study approach. *Information Processing & Management*, *30*(2), 205–221. https://doi.org/10.1016/0306-4573(94)90065-5

Sohn, T., Li, K. A., Griswold, W. G., & Hollan, J. D. (2008). A diary study of mobile information needs. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* (pp. 433–442). New York, USA: ACM Press. https://doi.org/10.1145/1357054.1357125

Sormunen, E. (2002). Liberal Relevance Criteria of TREC – Counting on Negligible Documents? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 324–330). New York, USA: ACM Press. https://doi.org/10.1145/564376.564433

Sridhar, M. S. (1995). Understanding the User - Why , What and How ? *Library Science with a Slant to Documentation and Information Studies*, *32*(4), 151–164.

Statista. (2018). Number of daily active Snapchat users from 1st quarter 2014 to 1st quarter 2018 (in millions). Retrieved from

https://www.statista.com/statistics/545967/snapchat-app-dau/

Stebbins, R. A. (1977). The Amateur: Two Sociological Definitions. *The Pacific Sociological Review*, *20*(4), 582–606.

Stebbins, R. A. (1980). "Amateur" and "Hobbyist" as Concepts for the Study of Leisure Problems. *Social Problems*, *27*(4), 413–417. https://doi.org/10.1525/sp.1980.27.4.03a00040

Stebbins, R. A. (1982). Serious Leisure: A Conceptual Statement. *The Pacific Sociological Review*, *25*(2), 251–272.

Stebbins, R. A. (1997). Casual leisure: a conceptual statement. *Leisure Studies*, *16*(1), 17–25. https://doi.org/10.1080/026143697375485

Stebbins, R. A. (2004). Pleasurable Aerobic Activity: A Type of Casual Leisure with Salubrious Implications. *World Leisure Journal*, *46*(4), 55–58. https://doi.org/10.1080/04419057.2004.9674374

Stebbins, R. A. (2005). Project-based leisure: theoretical neglect of a common use of free time. *Leisure Studies*, *24*(1), 1–11. https://doi.org/10.1080/0261436042000180832

Stebbins, R. A. (2007a). A Leisure-Based, Theoretic Typology of Volunteers and Volunteering. LSA newsletters No. 78.

Stebbins, R. A. (2007b). *Serious Leisure: A Perspective for Our Time*. New Brunswick, N.J.: Transaction Publishers.

Stebbins, R. A. (2009). Leisure and Its Relationship to Library and: Information Science: Bridging the Gap. *Library Trends*, *57*(4), 618–631. https://doi.org/10.1353/lib.0.0064

Stebbins, R. A. (2017). *Leisure's Legacy: Challenging the Common Sense View of Free Time*. Cham, Switzerland: Springer International Publishing.

Tang, R., & Solomon, P. (2001). Use of relevance criteria across stages of document evaluation: On the complementarity of experimental and naturalistic studies. *Journal of the American Society for Information Science and Technology*, *52*(8), 676–685. https://doi.org/10.1002/asi.1116

Taylor, A. (2009). *Relevance criterion choices in relation to search progress*. Rutgers, The State University of New Jersey. https://doi.org/10.7282/T3W959FQ

Taylor, A. (2013). Examination of work task and criteria choices for the relevance judgment process. *Journal of Documentation*, *69*(4), 523–544. https://doi.org/10.1108/JD-12-2011-0054

Taylor, A., Zhang, X., & Amadio, W. J. (2009). Examination of relevance criteria choices and the information search process. *Journal of Documentation*, *65*, 719–744. https://doi.org/10.1108/00220410910983083

Taylor, S. J., Bogdan, R., & DeVault, M. L. (2016). *Introduction to qualitative research methods : a guidebook and resource* (4th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Teijlingen, E. R. van, & Hundley, V. (2001). The importance of pilot studies. *Social Rsearch Update*, *35*.

Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess Web pages for information seeking. *Journal of the American Society for Information Science and Technology*, *56*(4), 327–344. https://doi.org/10.1002/asi.20106

Trattner, C., & Elsweiler, D. (2017). Food Recommender Systems: Important Contributions, Challenges and Future Research Directions.

Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, *56*(5), 540–562. https://doi.org/10.1108/EUM0000000007127

Voorhees, E. M. (2000). Variations in relevance judgments and the measurement of retrieval effectiveness. *Information Processing & Management*, *36*(5), 697–716. https://doi.org/10.1016/S0306-4573(00)00010-8

Wakeling, S., Halvey, M., Villa, R., & Hasler, L. (2016). A Comparison of Primary and Secondary Relevance Judgements for Real-Life Topics. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval - CHIIR '16* (pp. 173–182). New York, USA: ACM Press. https://doi.org/10.1145/2854946.2854968

Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, *49*(2), 115–133. https://doi.org/10.1002/(SICI)1097-4571(1998)49:2<115::AID-ASI3>3.0.CO;2-1

Wang, P., & White, M. D. (1999). A Cognitive Model of Document Use during a Research Project. Study II. Decisions at the Reading and Citing Stages. *Journal of the American Society for Information Science*, *50*(2), 89–114. https://doi.org/10.1002/(SICI)1097-4571(1999)50:2<98::AID-ASI2>3.0.CO;2-L

Wen, L., Ruthven, I., & Borlund, P. (2006). The effects on topic familiarity on online search behaviour and use of relevance criteria. In *Advances in Information Retrieval: Proceedings of the 28th European Conference in Information Retrieval (ECIR 2006). Lecture Notes in Computer Science* (Vol. 3936 LNCS, pp. 456–459). Springer. https://doi.org/10.1007/11735106_40

Westman, S., & Oittinen, P. (2006). Image retrieval by end-users and intermediaries in a journalistic work context. In *Proceedings of the 1st international conference on Information interaction in context - IIiX* (pp. 102–110). New York, NY, USA: ACM. https://doi.org/10.1145/1164820.1164843

Wilson, M. L., Alhodaithi, B., & Hurst, M. (2012). A diary study of information needs

produced in casual-leisure reading situations. *CEUR Workshop Proceedings*, *836*, 25–28.

Wu, H.-K., Krajcik, J. S., & Soloway, E. (2001). Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching*, *38*(7), 821–842. https://doi.org/10.1002/tea.1033

Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2012). A diary study on the happy worker: How job resources relate to positive emotions and personal resources. *European Journal of Work and Organizational Psychology*, *21*(4), 489–517. https://doi.org/10.1080/1359432X.2011.584386

Xie, I., & Benoit, E. (2013). Search result list evaluation versus document evaluation: similarities and differences. *Journal of Documentation*, *69*(1), 49–80. https://doi.org/10.1108/00220411311295324

Xu, Y. (2007). Relevance Judgment in Epistemic and Hedonic Information Searches. *Journal of the American Society for Information Science and Technology*, *58*(2), 179–189. https://doi.org/10.1002/asi.20461

Xu, Y. (Calvin), & Chen, Z. (2006). Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology*, *57*(7), 961–973. https://doi.org/10.1002/asi.20361

Yang, C., Hsu, Y.-C., & Tan, S. (2010). Predicting the determinants of users' intentions for using YouTube to share video: moderating gender effects. *Cyberpsychology, Behavior and Social Networking*, *13*(2), 141–152.

Yang, M. (2005). *An exploration of users' video relevance criteria*. University of North Carolina at Chapel Hill.

Yang, M., & Marchionini, G. (2005). Exploring users' video relevance criteria-A pilot study. In *Proceedings of the American Society for Information Science and Technology* (Vol. 41, pp. 229–238). https://doi.org/10.1002/meet.1450410127

Ye, C., & Wilson, M. L. (2014). The Characteristics of Casual Sessions in Search Behaviour Logs. In *Searching4Fun workshop at IIiX2014 Proceedings of the 5th Information Interaction in Context Symposium* (pp. 4–7). New York, NY, USA: ACM.

Yeh, N. (2016). The Role of Online Videos in Undergraduate Casual-Leisure Information Behaviors. *Global Journal of Human Social Science: C Sociology & Culture*, *16*(5), 7–13.

YouTube. (2018). YouTube for Press. Retrieved June 17, 2018, from https://www.youtube.com/yt/press/en-GB/statistics.html

Appendix A

Materials for the diary study

Appendix A.1 Initial diary design

Appendix A.2 Modified diary design

Appendix A.3 Final diary designs

Appendix A.4 Consent form

Appendix A.5 Information sheet

Appendix A.6 Demographic information form

Appendix A.7 Flyer

Appendix A.8 Invitation email

Appendix A.9 Email sent to potential participant

Appendix A.10 Receipt

Appendix A.11 Participant's feedback questions

Appendix A.12 Comparing culture and gender differences

Appendix A.1

Initial diary design (version 1)

Instructions:

We would like you to make an entry each time you look for videos on YouTube during your leisure time. The reasons for searching these videos should not be something related to work and they have to be personal. For instance, mothers who are looking for rhymes to their kids are not the type of tasks we interested in.

We are interested in the criteria you apply to judge the relevance of the retrieved videos. In other words, the reasons that make this video relevant or not relevant to your need. This could be the video title, the name of the channel, the date of upload or anything else.

Date **Time**

1. Describe the topic of your search session, what are you looking for? E.g. 'muffin recipe', 'makeup tutorial', 'historical documentary' or any other topic.

Or in case you are not looking for videos on a specific topic, please provide the motivation of this search, for example, 'kill the time', 'change my mood' or any other motivation.

---

2. Your location while doing this search

☐Home ☐work ☐ coffee shop ☐others please specify ____

3. Please do the following for **five** videos you chose to watch from the search result list:

a) Identify the relevance of this video to your need (relevant, partially relevant or not relevant)

b) Describe the reasons that contributed to your decision e.g. "I am looking for recent videos and the video was uploaded two days ago, that make it interesting".

Use the following table to fill the required information.

Video	Relevance judgment	Criteria applied
Video #1	☐Relevant ☐Partially relevant ☐Not relevant	
Video #2	☐Relevant ☐Partially relevant ☐Not relevant	
Video #3	☐Relevant ☐Partially relevant ☐Not relevant	
Video #4	☐Relevant ☐Partially relevant ☐Not relevant	
Video #5	☐Relevant ☐Partially relevant ☐Not relevant	

Appendix A.2

Modified Diary design (version 2)

Instructions

We would like you to make an entry each time you look for videos on YouTube during your leisure time. The reasons for searching these videos should not be something related to work and they have to be personal. For instance, students who are searching YouTube for academic lessons or mothers who are looking for rhymes to their kids are not the type of tasks we are interested in.

We are interested in the reasons that make you like the retrieved videos. This could be the video title, the name of the channel, the date of upload or anything else.

For each search session (each time you search YouTube for your need) fill one of the following pages. You could record up to ten sessions. The information required for each search session is as follows:

1. Date and time of your search session.
2. Describe the topic of your search session, what are you looking for? E.g. 'muffin recipe', 'makeup tutorial', 'historical documentary' or any other topic. Or in case you are not looking for videos on a specific topic, please provide the motivation of this search, for example, 'kill the time', 'change my mood' or any other motivation.
3. The location you are in while doing this search.
4. For three videos you found them useful, provide titles of these videos and describe why you liked them, e.g. "I am looking for recent videos and this video was uploaded two days ago, that makes it interesting".

Search Session # 1	
Date	**Time**
Click arrow, then use calender.	Enter time for this search session.
Topic / Motivation of your search	**Location** e.g. home, work, coffee shop
Click here to enter text.	Click here to enter text.
For <u>three</u> videos you liked, please write the title of the videos and reasons why you like them.	
Video #1 Title	**Reasons**
Click here to enter text.	Click here to enter text.
Video #2 Title	**Reasons**
Click here to enter text.	Click here to enter text.
Video #3 Title	**Reasons**
Click here to enter text.	Click here to enter text.
If there are videos that you did not complete them until the end and you decided to stop watching in the middle, please give the reasons for stopped watching them.	
Click here to enter text.	

Appendix A.3

Final Diary design (version 3)

My YouTube Diary

Instructions

We would like you to make an entry each time you search for videos on YouTube for **leisure or entertainment**. The reasons for searching these videos should not be something related to work and they have to be personal. For instance, students who are searching YouTube for academic lessons or mothers who are looking for rhymes to their kids are not the type of tasks we are interested in.

More specifically, the main goal of this study is to know what criteria people use when deciding on the usefulness of the retrieved videos. In other words, we are interested in the reasons that make you like/dislike the retrieved videos. Those could be the novelty of the video content, title, the name of the channel or anything else.

For each search session fill one of the following pages. By search session we mean each time you search YouTube for your need, you could fill more than one form a day if you search YouTube several times that day. You could record up to ten sessions. The information required for each search session is as follows:

1. Date and the time of the day (morning, afternoon, evening).
2. Describe the topic of your search session, what are you looking for? E.g. 'muffin recipe', 'makeup tutorial', 'historical documentary' or any other topic.

   In case you are not looking for videos on a specific topic, please move to the next field to provide the motivation of this search, for example, 'kill the time', 'change my mood' or any other motivation.

3. The location you are in while doing this search
4. The device used (desktop, tablet, mobile phone).
5. For three videos you chose to watch, provide titles of these videos and describe why you decided to click on them, e.g. "I am looking for recent videos and this video was uploaded two days ago, that makes it interesting".

6. In case you found the selected video not useful and decided to stop watching it, please provide the reasons.

Please keep recording your search sessions in this document for a duration of one week and make sure that you record the information while you (or soon after) searching YouTube to avoid any memory lapses.

تعليمات

نود منك تسجيل بعض البيانات في كل مرة تبحث/تبحثين عن مقاطع فيديو باستخدام برنامج اليوتيوب لأغراض الترفيه. تهتم هذه الدراسة بعمليات البحث الشخصية والتي لا تتعلق بالعمل أو الدراسة. على سبيل المثال, عمليات البحث التي يقوم بها الطلاب للحصول على دروس أكاديمية او الأمهات اللاتي يبحثن عن أناشيد لأطفالهم ليست من اهتمام هذه الدراسة.

الهدف الرئيسي من هذه الدراسة هو معرفة المعايير التي يقيم المستخدم على أساسها مدى ملائمة مقاطع الفيديو المستردة لحاجته. بمعنى آخر نحن مهتمون بالأسباب التي تجعل المستخدم يحب /أو لا يحب مقاطع الفيديو التي قام باستردادها. هذه الأسباب قد تكون لأن محتوى الفيديو لم يسبق للمستخدم مشاهدته من قبل, عنوان الفيديو جذاب, ثقة المستخدم باسم قناة اليوتيوب التي عرضت المقطع أو لأي سبب آخر.

لكل جلسة بحث الرجاء تعبئة واحده من الصفحات التالية. بجلسة البحث نعني كل مره قضيت فيها وقت وأنت تبحث/تبحثين في موقع اليوتيوب عن حاجتك, بامكانك تعبئة أكثر من صفحة باليوم اذا كنت قمت بالبحث لأكثر من مره في هذا اليوم. يمكنك تعبئة عشر استبانات كحد أقصى (ليس بالضروره عشر استبانات, الرجاء تعبئة الاستبانات بعدد جلسات البحث التي تقومين بها). المعلومات المطلوبة في كل جلسة بحث هي:

١. التاريخ والوقت.
٢. وصف الموضوع الذي تبحثين عنه, عن ماذا تبحث؟ مثال: وصفة كعك, فلم وثائقي أو أي موضوع آخر. في حال كنت لا تبحث عن موضوع معين, الرجاء الانتقال للحقل التالي لذكر الدافع لهذا البحث, على سبيل المثال, "قضاء وقت" , "تغير حالتي المزاجية" أو أي دافع للبحث في يوتيوب.
٣. الموقع الذي تتواجد فيه عند قيامك بالبحث.
٤. الجهاز المستخدم (كمبيوتر مكتبي, جهاز لوحي, هاتف متنقل "جوال").
٥. لثلاثة مقاطع فيديو اخترت مشاهدتها, الرجاء تزويدنا بعناوينها ووصف الأسباب التي جعلتك تقرر الضغط على روابطها من قائمة نتائج البحث. مثال: "أنا أبحث عن مقاطع حديثة, وهذا المقطع تم تحميله من يومين".
٦. في حال وجدت المقطع الذي اخترت مشاهدته غير مجدي وقررت التوقف عن مشاهدته, الرجاء ذكر الأسباب.

لطفا, استمر/استمري بتقيد جلسات بحثك على موقع اليوتيوب في هذا الملف لمده أسبوع واحد فقط وتأكد من تقيد المعلومات في هذا الملف في نفس وقت مشاهدتك لليوتيوب (أو بعده بوقت قصير) حتى تكون المعلومات دقيقة وتتجنب نسيانها.

Search Session #1			
Date	**Time**(morning, afternoon, evening)	**Location** e.g. home, work	**Device used** e.g. desktop, tablet, mobile
What are you looking for? e.g. "muffin recipe" **note:** In case you are not looking for a certain topic skip this field		**Why you start this search?** e.g. kill the time, change my mood, I want a new dish for the dinner	
For <u>three</u> videos you choose to watch, please write the title of the videos and the reasons that make you select these videos			
Video # 1 Title		**Reasons**	
If you stopped watching this video and did not complete it until the end, please give the reasons for that.			
Video # 2 Title		**Reasons**	
If you stopped watching this video and did not complete it until the end, please give the reasons for that.			
Video # 3 Title		**Reasons**	
If you stopped watching this video and did not complete it until the end, please give the reasons for that.			

Appendix A.4

Consent Form (Diary study)

- I confirm that I have read and understood the information sheet for the above project and the researcher has answered any queries to my satisfaction.
- I understand that my participation is voluntary and that I am free to withdraw from the project at any time, up to the point of completion, without having to give a reason and without any consequences.
- I understand that I can withdraw from the study any personal data (i.e. data which identify me personally) at any time.
- I understand that anonymised data (i.e. .data which do not identify me personally) cannot be withdrawn once they have been included in the study.
- I understand that any information recorded in the investigation will remain confidential and no information that identifies me will be made publicly available.
- I consent to being a participant in the project.
- I consent to being audio recorded as part of the project.

I agree to take part in this study

Appendix A.5

Information Sheet (Diary study)

Thank you for your interest in this study. My name is Sarah Albassam and I am a PhD student at the Department of Computer and Information Science, University of Strathclyde. I am conducting research on relevance criteria and require volunteer participants. This study is supervised by Prof Ian Ruthven and has been granted ethical approval from the Department of Computer and Information science ethics committee.

What is the purpose of this investigation?

The main goal of this study is to investigate how adults decide about the usefulness of retrieved videos in their leisure time. We are interested in the criteria users apply in making relevance judgment decisions when searching videos for leisure or entertainment. Investigating relevance criteria is essential to gain more understanding of the user judgment behaviour and to aid in designing more useful IR systems.

Do you have to take part? Your participation in this study is voluntary. If you do agree to participate, you can withdraw from the participation at any time during the study without comments or penalty.

What will you do in the project?

Your participation will involve the maintenance of a diary (printed or electronic based on your preferences) for a one week. In this diary, you will be asked to make an entry each time you search for videos on YouTube for leisure or entertainment. The reasons for searching these videos should not be something related to work and they have to be personal. For instance, students who are searching YouTube for academic lessons or mothers who are looking for rhymes to their kids are not the type of tasks we are interested in. If you are happy to be contacted after the diary, a follow-up short informal interview will be held in order to go through your diary entries and clarify any ambiguity. As an appreciation for your time and effort you have made in the diary, a £ 5 voucher will be offered to you.

Why have you been invited to take part?

People who search YouTube more than twice a week for leisure or entertainment purposes are welcome to participate.

What are the potential risks to you in taking part?

There are no risks for participating in this study.

What happens to the information in the project?

All the diary entries will be kept entirely confidential. All the data will be securely stored on password protected computers and paper data will be stored securely in the Livingston Tower building at the University of Strathclyde. The results of this study will be used for research purposes only.

The University of Strathclyde is registered with the Information Commissioner's Office who implements the Data Protection Act1998. All personal data on participants will be processed in accordance with the provisions of the Data Protection Act 1998.

Thank you for reading this information – please ask any questions if you are unsure about what is written here.

What happens next?

If you are happy to take part in this study please indicate this in the consent form below then press the arrow button to continue. If you do not wish to be involved in this study, we would like to thank you for your attention. For more information please do not hesitate to contact me.

Researcher Contact Details:
Sarah Albassam
PhD student
Department of Computer and Information Science
University of Strathclyde, Livingstone Tower, room 1214
Email: sarah.albassam@strath.ac.uk

Supervisor Contact Details:
Prof Ian Ruthven
Department of Computer and Information Science
University of Strathclyde, Livingstone Tower
Email: ian.ruthven@strath.ac.uk

This investigation was granted ethical approval by the Computer and Information Science Department Ethics Committee.

Appendix A.6

YouTube Diary study - Demographic Information Form

Age:

Gender:

- Male
- Female

Profession

Specialization

How often do you search for videos on YouTube for leisure or entertainment?

- Rarely
- 1-2 times a month
- 1-2 times a week
- Three or more times a week
- Once a day
- Several times a day

Contact Email address:

Appendix A.7

Diary Flyer

Earn £5

Department of Computer and Information Sciences

University of Strathclyde Science

For contributing to **a YouTube diary study** that aims to understand how people decide about the usefulness of the videos they selected to watch for leisure or entertainment.

We are interested in the reasons that make you like/dislike your retrieved videos.

You are welcomed to participate in your own time and place, and in your own choice of formats- printed document, electronic word document or a shared google document.

As an appreciation of your time and effort, an amount of **£ 5 shopping voucher** will be offered to all participants.

For more information or If you would like to take part in the study, please contact:

Sarah Albassam- sarah.albassam@strath.ac.uk

Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk

Appendix A.8

Invitation email

Dear all

As part of my research project, I am looking for participants who are willing to maintain a short YouTube diary for a period of one week. The goal of the study is to understand how people decide about the usefulness of the videos they selected to watch for leisure or entertainment.

We are interested in the reasons that make you like/dislike your retrieved videos.

The required task is to simply search YouTube at your free time as you do usually and provide the reasons that make you select your retrieved videos.

If you are interested in participating or just curious to learn more, please do not hesitate to contact me

Sarah Albassam

E-mail: sarah.albassam@strath.ac.uk

Appendix A.9

Email sent to potential participants

Dear ……

Thank you for your interest in this study. The main goal of the study is to understand how YouTube users select which video to watch among many other videos returned from their search in the search result list. I am interested in the reasons that make you choose the one you choose. In other words your criteria in selecting your videos.

If you are interested, please fill the following form which collects few demographic information (e.g. age, gender) and contains the consent form

https://strathsci.qualtrics.com/SE/?SID=SV_8ieQUND9omnil0l

Then use this google document or the attached word document to record your inputs.

goo.gl/xBfZOD

The Google document contains instructions and clarification on how to fill the diary. Please read them carefully and feel free to contact me if you have any questions.

To clarify the required task, I attached a sample diary that contains three examples. This sample is just an example of what participants could record, however, the reasons for like/dislike your retrieved videos are not limited to the reasons shown in the examples. The examples were supplied purely to get participants thinking about the kinds of reasons they could record and to show the level of and type of details expected.

I would like to clarify some more points:

- Please make sure that you record the information while you (or soon after) searching YouTube to avoid any memory lapse and keep recording your search sessions for one week.
- You do not have to complete all the 10 sessions (pages) during the one week period, just record sessions as you do normally.
- For each time (search session) you searched YouTube it is not necessary to watch 3 videos, it could be one or two videos
- You also can watch videos in any language, but fill the diary in English.

Many Thanks

Sarah

Appendix A.10

Receipt-Diary Study

I have received a £5 shopping voucher for participating in a YouTube Diary study, which is part of a PhD research for Sarah Albassam at the Department of Computer and Information Science - University of Strathclyde.

Name

Date

Signature

Appendix A.11

Participants' feedback questions

1. How was your experience in filling out this diary?
2. Do you think I have asked the right questions? Or I might design it better?
3. Are the search sessions provided in your diary represent the typical way of how you use YouTube?

Appendix A.12

Comparing culture and gender differences

Criteria	KSA participants	UK participants	Females	Males
coverage	✓	✓	✓	✓
topicality	✓	✓	✓	✓
recency	✓	✓	✓	✓
genre	✓	✓	✓	✓
length	✓	✓	✓	✓
people in the video	✓	✓	✓	✓
background experience or personal memories	✓	✓	✓	✓
novelty	✓	✓	✓	✓
familiarity	✓	✓	✓	✓
affectiveness	✓	✓	✓	✓
serendipity/ curiosity	✓	✓	✓	✓
habit	✓	✓	✓	✓
time constraint	✓	✓	✓	✓
quality of source	✓	✓	✓	✓
content quality	✓	✓	✓	✓
technical quality	✓	✓	✓	✓
cinematography	✓	✓	✓	✓
visual appeal	✓	✓	✓	✓
sound/voice	✓	✓	✓	✓
cost		✓	✓	
language/subtitle	✓	✓	✓	
version	✓	✓	✓	✓
availability	✓	✓	✓	✓
verification	✓	✓	✓	✓
unusualness	✓	✓	✓	✓
rank order	✓	✓	✓	✓
popularity	✓	✓	✓	✓
recommended video	✓	✓	✓	✓

Appendix B

Materials for the interviews study

Appendix B.1 Flyer

Appendix B.2 Invitation email

Appendix B.3 Instruction sheet

Appendix B.4 Consent form

Appendix B.5 Search scenario

Appendix B.6 Demographic form

Appendix B.7 Pre-search interview questions

Appendix B.8 Receipt

Appendix B.1

Earn £5

Department of Computer and Information Sciences

For contributing to **a YouTube study** where you search as you do normally followed by an informal discussion. The study aims to understand how people decide about the usefulness of the videos they selected to watch for leisure or entertainment.

As an appreciation of your time and effort, a £ 5 will be offered to all participants.

For more information or If you would like to take part in the study, please contact:
Sarah Albassam- sarah.albassam@strath.ac.uk

Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk	Sarah.albassam@strath.ac.uk

Appendix B.2

Invitation email

Hi

As part of my research, I am looking for YouTube users to take part in a **one hour** search and discussion session. The main goal of the study is to understand how YouTube users decide about the usefulness of the videos they selected to watch for leisure or entertainment.

I would kindly invite you to Livingston Tower level 12 room 1214 where you can search and view videos on YouTube as you normally do in your free time. A short discussion will follow the search session.

As an appreciation of your time and effort, a **£ 5 shopping voucher** will be offered to all participants.

If you are interested, kindly contact me on my email (sarah.albassam@strath.ac.uk) so we could agree on a suitable time to meet and chat. Your help would be much appreciated.

Thanks in advance

Sarah

Appendix B.3

Instructions sheet (second study)

Thanks again for coming today and agreeing to participate in this study. My name is Sarah Albassam and I am a PhD student at the Department of Computer and Information Science. The main goal of the study is to understand how YouTube users decide about the usefulness of the videos they selected to watch for leisure or entertainment.

I am interested in the reasons that make you choose the one you choose. In other words your criteria in selecting your videos. Also, what information elements such as date, title were important to you. In order to achieve this goal, I would kindly ask you to conduct a search using YouTube as you do normally.

Your YouTube search session will be recorded. The recordings will help in understanding more about how you select your videos. **I am not judging what you watch**, just want to understand the video viewing process which will help in improving the video systems. So, take this search session as a relaxing activity.

At the end of the search session, we will replay the recording and discuss points where you did some relevance judgments (videos selection and viewing). All your entries will keep entirely confidential and the results of this study will be used for research purposes only.

If you are happy to take part in this study I would kindly ask you to sign a consent form to confirm your agreement to participate. And to fill a short demographic information form.

Do you have any questions?

Appendix B.4

Consent Form

- I confirm that the researcher has explained the goal and process of the study and answered any queries to my satisfaction.
- I understand that no personal judgment will be made on the data collected.
- I understand that my participation is voluntary and that I am free to withdraw from the study at any time, up to the point of completion, without having to give a reason and without any consequences.
- I understand that I can withdraw from the study any personal data (i.e. data which identify me personally) at any time.
- I understand that anonymised data (i.e. .data which do not identify me personally) cannot be withdrawn once they have been included in the study.
- I understand that any information recorded in the investigation will remain confidential and no information that identifies me will be made publicly available.
- I consent to be a participant in the study.
- I consent to be audio recorded as part of the study.
- I consent to have my search session video recorded.

(PRINT NAME)	
Signature of Participant:	Date:

Appendix B.5

Search scenario

Imagine you got 20 minutes spare time and you decided to watch some videos during this time to entertain yourself. You might recall some of your recent searching or browsing for videos on YouTube that you did for **leisure or entertainment** and reply it. The search topic should be personal **and not related to any course assignments.** There is no restriction on how you initiate your searches. E.g. starting by typing query, YouTube suggestions and popular videos, or your subscriptions channels.

Appendix B.6

Demographic form

Gender:

- ○ Male
- ◉ Female

Major course of study

Are you

- ○ Undergraduate student
- ○ Master Student
- ○ PhD student
- ○ Other

Age:

How often do you search for videos on YouTube for leisure or entertainment?

- ○ Rarely
- ○ 1-2 times a month
- ○ 1-2 times a week
- ○ Three or more times a week
- ○ Several times a day

Appendix B.7

Pre-search questions

1. In your free time, what usually motivates you to view videos on YouTube? (E.g. pass time, change mood learning)

2. How you use YouTube for entertainment? What are your favourite videos? Video content that interest you?

3. How do you usually initiate your search? Browsing trending videos, check channels that you subscribe to, search by typing queries?

Appendix B.8

Receipt-Interview

I confirm that I received the amount of five pounds (£5) from Sarah Albassam for my participation in her study.

Name

Date

Signature