Users' relevance criteria for video in leisure contexts

Sarah Albassam  Ian Ruthven

University of Strathclyde
Users' Relevance Criteria for Video in Leisure Contexts

Abstract

Purpose - The purpose of this paper 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?

Design/methodology/approach - A naturalistic diary was performed in which 30 participants completed diaries, providing details on their video relevance criteria. The analysis revealed 28 relevance criteria grouped into eight categories.

Findings - Twenty-eight relevance criteria were identified through the analyses of the diaries’ content and they were grouped into 8 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, 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.

Research limitation/implication - The decision to follow a naturalistic approach reduced the level of control on the study. A further limitation can be found in the participants' sample used in this study, all the participants of the main study were university or college students.

Practical implication - This study attempted to enrich the current literature by investigating users’ video relevance criteria in leisure contexts. This investigation might have implications on the design of video search systems.

Originality/value - Previous relevance criteria studies focused on work contexts and the information judged was mainly in text format. This paper outlines new insights by investigating video relevance criteria in leisure context.

Keywords Relevance criteria, Casual leisure, Naturalistic diary study, Video retrieval

Paper type Research paper
1. Introduction

Relevance is a core concept in Information Retrieval. The ultimate goal of all Information Retrieval (IR) systems is to retrieve relevant documents that satisfy user information needs. Many studies have attempted to understand how people judge the relevancy of retrieved documents resulting in an enormous amount of work concerning user relevance judgment, behavior 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. In other words, what are the relevance criteria that users apply when making a relevance judgment decision? Although there is a rich literature on different aspects of relevance criteria, there are still open questions. For example, there is a lack of knowledge regarding how users apply relevance criteria in video relevance judgments. A few studies have focused on video relevance, such as Yang (2005), however, these studies are limited to work-related tasks.

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). Little, though is known about the relevance criteria that are important to users in leisure contexts and how these might be different from other contexts. Thus, we need research to fill in the gap of knowledge regarding video relevance criteria in leisure context.

The importance of this study lies in its attempt to enrich the current literature by investigating users' video relevance criteria in leisure contexts. By shedding light on this uncovered area of research, the IR community will gain a deeper understanding of how users make their relevance judgment decisions in a leisure context. Furthermore, the findings of this study will have implications 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. In addition, the design of video retrieval systems may also be different from text retrieval systems. 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 leisure contexts will help in designing systems that address their needs.

This study was conducted to investigate the relevance criteria users apply when judging 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?
2. Related Work

2.1 Relevance Criteria Studies

Early research has significantly contributed to the user relevance criteria literature and has identified key criteria and synthesized them into different categories. Schamber (1991) mentioned 22 criteria classified into 10 categories gathered from 30 occupational users of weather information. In a similar work, Barry (1994) identified 23 criteria grouped into seven categories from faculty and student users. Later, Barry and Schamber (1998) combined and compared the results of their previous studies. The comparison yielded 10 criteria in common. The criteria are: depth/scope/specificity, accuracy/validity, clarity, currency, tangibility, quality of sources, accessibility, availability of information/sources of information based on their research, verification, and affectiveness.

Based on Barry and Schamber (1998), other studies emerged to cover different aspects of relevance criteria. For example, group of studies in relevance criteria literature (Wen, Ruthven & Borlund (2006); Tombros, Ruthven & Jose (2005); Balatsoukas & Ruthven (2012)) have focused on the factors that guide the relevance criteria selection process. Topic familiarity, searching task type and the ranking order of the result list are some of the factors that found to have an effect on the users' relevance criteria choices.

The findings of relevance criteria 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 and video retrieval. Laplante (2010) explored relevance judgments in the context of music retrieval for leisure purposes. The findings indicated that in the context of music retrieval, criteria related to the participant's beliefs and preferences have a significant influence on relevance judgment decision and topicality was not mentioned by the study participants. Choi and Rasmussen (2002) attempted to discover the relevance criteria that users employ when making relevance judgments of images. They reported that topicality dominates among other criteria. In addition, other criteria like image quality and clarity seems to be important to users. They also noticed differences in the priority of applying the criteria at different search stages with the importance of criteria such as appeal of information and accessibility increased as the search process progress whilst the importance of topicality decreases.

There were also limited studies that investigated relevance criteria in video retrieval. In her PhD study, Yang (2005) explored the criteria people applied when making video relevance judgments.
New criteria such as cinematography and scene-level information were identified with topicality remaining the most dominant criterion.

Another attempt to explore everyday video seeking behavior and video information needs is the work of Cunningham and Nichols (2008). This study mentions briefly some elements of the video retrieval system that are used in relevance judgment decision such as comments, thumbnails and number of views. However, it says nothing about other aspects of personal criteria such as novelty and authority. Moreover, the study focused on the behavior of locating videos in everyday life rather than decisions of selecting and watching videos (which is our study focus).

2.2 Leisure
Stebbins pioneered in proposing the serious leisure perspective (SLP), a theoretic framework that bridges and synthesizes three forms of leisure: serious, casual and project-based. 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).

Leisure is worth investigating in order to examine whether the qualities of information change when moving from task-based scenarios in work contexts to leisure contexts. Later, researchers in the information behavior and seeking field start to draw more attention towards contexts other than professional and academic. Hartel, as an example, investigated the information needs and seeking aspects of serious leisure activities such as cooking gourmet (Hartel, 2006).

Elsweiler, Wilson & Lunn (2011) explored the information behavior of two different casual leisure scenarios. The first focuses on the information behavior relating to use of television and the second was the casual information behavior of Twitter users. The result revealed variety of needs with different level of complexities. It also found that a lot of motivations provided by the participants are less dependent on finding certain information, the experience itself was the main concern of the participants. Similar findings were reported in the second study.
More similar to our study, Yeh (2016) explored casual-leisure information behaviors when searching video online. Data was collected using diaries followed by semi-structured and in depth interviews. The study proposed a framework of casual-leisure video viewing processes and information behaviors. The framework divided the process of viewing videos online into three phases according to time: pre-viewing, viewing, and post-viewing. Information behaviors in each phase were included in the framework. Although the main goal of the casual-leisure search is to satisfy hedonic needs, the videos viewed in this context trigger information needs that encourage the participant to search for further relevant information. While Yeh’s study share similarities with our study in terms of the methods and the context (casual leisure video viewing), the focus is different. Yeh’s study investigated the motivations that trigger casual leisure video search and the information behavior activities while viewing videos. In contrast, our study focuses on the relevance criteria applied when judging videos in leisure context.

2.3 YouTube
According to Cisco, video content will dominate the internet in the coming years. In particular, video traffic acquired 73 percent of all consumer internet traffic in 2016 and this percent is expected to be 82 by 2020 (Cisco, 2017). Among different video search engines, YouTube is the most popular and widely use site which is ranked second globally after Google. (Alexa, 2017). Thus, studies emerged to discover different aspects of users' behaviors on YouTube. For example, Madden, Ruthven and McMenemy (2012) conducted a content analysis to 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 behaviors when searching videos on YouTube with the ultimate goal of comparing these behaviors to text searching. To achieve this goal, a sample of YouTube videos were built using a web crawler. The findings showed that users intend to benefit and 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 focused 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 behaviors for news content on YouTube. In specific, the study focused on the factors influencing viewing and sharing news videos and the motives that stimulate such activities. The findings revealed different motives for different types of news-related content. Yang, Hsu and Tan (2010) examined users' motivations for sharing their own videos on YouTube. The authors went one step further by
examining gender differences in sharing behaviors. They found that female users’ intention is related to usefulness and social norms, while male users’ intention is influenced by interpersonal norms.

2.4 Literature Summary

Previous related studies in leisure or YouTube contexts explored the information seeking of people with more focus on the users’ needs and motivations rather than relevance criteria. Furthermore, the relevance criteria literature clearly show how previous works in relevance criteria have focused on work-related tasks and more specifically in academic or professional context. In addition, the vast body of relevance criteria literature were on searching for textual information rather than audio or visual information. Therefore, our study attempts to shed the light on video relevance criteria in leisure contexts in order to fill in this knowledge gap.

3. Diary Study

3.1 Overview

With the aim of investigating video relevance criteria in leisure context, we selected a naturalistic diary method as the main data collection method. As leisure is known to be an everyday life activity, we needed an instrument which could collect data in a more natural and realistic way.

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 and Jones, 2007). Therefore, it is a good research instrument choice for research that aims to understand the "why" aspects of user behaviors, in our 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 (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 good commitment from the participants who should be convinced to complete the diaries (Pickard, 2013). Another issue is that participants might miss valuable input because they thought it is not sufficiently interesting to record it. Thus, the level of detail required has to be clarified to participants at the beginning of the study (Palen & Salzman, 2002).

To minimize the limitations of the diary method, we designed the diary in a way that would minimize the time and effort needed to complete it. A semi-structured diary was designed which enabled the participants to provide their inputs to pre-defined fields for a period of one week. Two pilot studies preceded the main diary study to ensure the validity of the diary design and improve it. To help the
participants understand the type of information they should record and the level of details, a sample of
diary entries were provided as part of the diary. The final version of the diary is shown in the
Appendix. This study was granted approval from the Departmental Ethics Committee.

3.2 Participants and Recruitment
Thirty YouTube users participated in the main diary study. Twenty of the participants were female
and 10 male. Participants were young adults between 17 and 35 with average age of 23. Among the
participants 18 were undergraduate students, 5 PhD students, 3 Masters students, 2 college students
and 2 unemployed graduate students. They were studying different disciplines at various universities.
In general, participants acknowledge themselves as heavy YouTube users. Twenty 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 8 participants (6 females and 2 males) from the previous pilot studies were also
included in the analysis. Participants were recruited through emails and flyers distributed at different
universities and college. The recruitment took place in the United Kingdom and Saudi Arabia. £5
shopping vouchers were offered to the participants.

3.3 Procedure
We first asked participants to complete a demographic form that collects information such as age,
gender, profession and the frequency of searching YouTube for leisure or entertainment. Then, a link
to a Google document was sent to each participant. Each participant had his own Google document to
record entries in, and this was shared with the first author only. We also offered a Word diary or
printed diary as an alternative to Google documents, 11 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 we had a live access to the participants’ diaries, so we could contact them when
something 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. 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. YouTube was chosen as the video
retrieval system in this study because it is the most popular system widely used by over a billion users
and with huge variety of videos. During the process, we kept in touch with the participants via email
to answer their questions, and encourage them to keep going with the work. After completing the
diaries we contacted the participants to set a time to meet and have a post-diary discussion where we
could clarify any ambiguities appeared in the diaries and to get the feedback of the participants and
provide them with the compensation.
4. Data Analysis

Preliminary analysis of participants’ diaries took place at early stages of the data collection process by
careful reading of the diaries. Each utterance in the ‘reasons’ fields of the diaries was considered as a
unit of analysis and assigned a code (relevance criterion). We split utterances that touched on more
than one relevance criterion. 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”).

The coding process followed an a priori coding approach because of the previous rich literature on
relevance criteria. Following this approach, coding schemes from previous relevance criteria studies
including (Balatsoukas & Ruthven, 2012; Barry, 1994; Barry & Schamber, 1998; Savolainen & Kari,
2006; Yang, 2005) were grouped together, and the selection of potential codes assigned to utterances
was based on this derived scheme. Codes or codes’ definitions from different studies that share
similarities with each other were compared and aggregated. For example, the Coverage criterion in
our study was based on the codes in Table 1 which were mentioned in previous studies.

“Insert Table # 1 here.”

Each utterance was assigned a preliminary code from the pre-mentioned relevance criteria list where
possible. The initial coding scheme derived from the literature was revised iteratively during the
coding of the data. 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 and Greenberg, 2006) which have an effect on relevance judgment in academic
context but did not apply in leisure context. Moreover, some of the criteria mentioned in previous
coding scheme have different labels but with the same meaning such as ‘genre’ and ‘resource type’;
‘layout’ and ‘presentation’; ‘audience’, ‘intended use’ and ‘orientation/level’. So we eliminate the
duplicated codes. Some codes were merged together into one criterion such as ‘reliability’ and
‘reputation’ are merged to form one criterion ‘quality of source’. During the coding process,
additional codes were added to the coding scheme as they emerges from the inductive content analysis
of the diaries. For example, ‘recommended video’, ‘People in the video’ and ‘habit’ were new codes
added to the scheme.

Finally, the criteria were clustered into categories based on similarity. The analysis process was
iterative between developing the coding scheme and assigning codes to participants' responses. Then
responses with the same code label were grouped together and checked again weather the code label
fit them properly. In many cases, we needed to examine the exact video on YouTube to understand
the content or discover names of YouTube channels, people or events.

In order to increase the reliability of the data analysis process, another coder were asked to code a
subset (10%) of the data chosen randomly. The Kappa coefficient of inter coder agreement of the two
coders found to be 0.70 which is considered as a good level of agreement.

5. Results
In total, the participants provided 234 search sessions with an average of six sessions per participant
(S.D. 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 (S.D. 8.2 min 2 max 30). The results reported in this section includes
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",
"Auto play" these responses were necessarily omitted from the analyses. After cleaning the data, a
total of 787 instances of relevance criteria was extracted from the diaries.

The participants most often watched YouTube on mobiles (87 sessions), laptops (71 sessions),
desktop (44 sessions) and tablets (seven sessions). The majority of leisure search sessions were
performed at home (180 sessions) the remaining were at work (21 sessions), university (16 sessions),
transportation e.g. bus and car (seven sessions), restaurants and coffee shops (five sessions) other
places such as a friend’s house, beauty salon and gym (three sessions). In terms of the time of the day,
leisure search sessions usually occurred in the evening time (105 sessions) followed by afternoon (71
sessions). The fewest number of sessions was in night time (55 sessions). There were some missing
data because of the cases where participants did not complete these fields.

5.1 Relevance Criteria
Twenty-eight relevance criteria were identified through the analyses of the diaries’ content and they
were grouped into eight categories as follow: (a) criteria related to the information content of the
video; (b) criteria related to the participants' previous experience and background; (c) criteria related
to the participant's beliefs and preferences or their situation; (d) criteria related to the quality aspects
of the video or the source providing the video; (e) criteria related to audio/visual features of the video;
(f) criteria related to the accessibility of the video; (g) criteria related to other information within the
environment; (h) criteria related to other people’s opinions or YouTube’s recommendations. These are
discussed 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.

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 such as "he's giving too much details". Moreover, responses related to the targeted audience of the video are also coded under depth code, for example "the title seemed appropriate for beginners". Responses related to the scope of the information include "I stopped because I got enough information". Specificity examples include "I did not saw it all because it's presented a lot of products and I am interesting in certain product, so I only saw the part that I am interesting in". Responses related to variety include "I chose this one in particular as it seemed to cover a fair range of topics". 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" counted in this category because the thumbnail gives a summary of the video.

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 our data. A typical response demonstrating the matching between the video's information and the participant's search topic is "That’s what I was searching for". 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". 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". In some cases, participants were predicting the video was going to match his/her interest such as "the title is interesting".

Recency is defined as the extent to which the video is recent and this is important to the user. Responses such as "uploaded recently", "looks 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.

Genre is defined as the extent to which the genre of the video (e.g. anime, historical, comedy) is a factor in the relevance judgment. An example of responses coded as Genre include "I like rap music".
Genre does not necessarily match the YouTube classification of the video which seems to be more general than Genre as it appears in our data.

**Length** is defined as the extent to which video length (duration) is a factor in the participant's judgment such as “Too long 10 min”. Another criterion, strongly related to Length is Time Constraint (see below).

**People in the video.** The extent to which the participant's judgment is influenced by people appearing in the video (TV host, singer, actor, band, YouTuber or guest etc.). For example, “David Dobrik is a funny guy”. In some responses, 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 judgment 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.

**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, such as “I chose this video knowing very little of the subject, in fact, I knew nothing at all about whatever the 'dark net' was”, “Memories of my own hamsters made me want to watch”. Background/experience means that 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.

**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 including “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”.

**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. Examples include “I just really like this
song”, “Uploaded by a channel I subscribe to, I chose this video because I always enjoy their videos”. 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.

*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” taking the criterion from (Savolainen and Kari, 2006). Responses that included mentions of emotions such as boring, cute, humor, laugh, funny, not interested, happy, exciting are assigned this code. Examples from the diaries include “Funny, and has a great sense of humor”, “I found the song boring, didn't match my taste in music”. This was the criterion with the second largest number of mentions in our data.

*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 accidental discovery of useful or interested 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”. 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”. In some cases, 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”. 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'”.

*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 selected familiar videos. What distinguishes it from Familiarity is the mention of repetitively watching the video For example, “I commonly watch this YouTuber and entertainment” or watching the video as a part of another habit “Continued to fill out some applications and enjoyed listening to music as I worked”.
**Time constraint.** Defined as the extent to which time constraint is a factor in participant’s judgment. This category is strongly related to Length. 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 include “I’d watch it till the end if I have time”. So the reason behind the relevance judgment decision is more than just the length of the video, it is mainly based on the participant situation and time that he could afford “I cannot afford to spend that much time”.

**Criteria related to the quality aspects of the video or the source providing the video.** This group of criteria is mainly focused on the video’s perceived quality, either content quality, technical quality or the quality of the source providing the video.

**Quality of Source.** Defined as 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. Examples of responses counted under this category include “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 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 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”.

**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”.

**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”. When the video has many versions and the participant choose 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 focused on the aesthetic characteristics of the audio/visual content rather than the quality aspects of the video.
**Cinematography** is defined as “the extent to which the video contained any specific film techniques the participant was interested in, such as camera movement, color, 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.).”

**Visual appeal** is defined as the extent to which the thumbnail was appealing to the participant such as “the thumbnail was colorful”. 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.

**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”, “have to watch the subtitles, because there is no talk only background music”. 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.

**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”.

**Language/subtitle** is defined as the extent to which the language that was spoken in the video is understandable by the participant, and if it was in a foreign language, whether there were subtitles shown in the video, e.g. “it is in German” and “have subtitles”.

**Version** is defined as the extent to which different versions exist and judgments 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”.

**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.

**Availability** is defined as the extent to which a number of videos that cover the same topic are available and judgments 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”.

*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”.

*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”.

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.

*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”.

*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 270K likes, so it might be good”. 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”.

*Recommended Video* is defined as the extent to which a participant's judgment 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 2 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 judgments.

“Insert Table #2 here.”

### 5.2 Topics and Motivations

Participants provided various search topics when answering the 'What are you looking for?' diary question. In many cases, the related 'Why 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 our study, 62% of leisure searches began with a specific information need (topic). The remaining search sessions lack the mention of specific topics and were triggered by various motivations. The main topics mentioned in the diaries are summarized in Table 3. The topics mentioned in our study overlap with the topics mentioned in previous studies of video retrieval (Chung & Yoon, 2012; Cunningham & Nichols, 2008; Lux, Lagger & Marques, 2011; Yeh, 2016).

“Insert Table #3 here.”

Similarly, the motivations of leisure searches is summarized in Table 4. The motivations for searching videos in leisure time mentioned in our study, are in line with previous studies. Cunningham and 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 subjects emotional state/ mood') and the examples provided matched examples classified under 'pass time', change mood' and 'entertainment' in our 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 our study, such responses were classified as 'recommendations' and 'other' respectively.

“Insert Table #4 here.”

The motivations that stimulated participants to initiate their searches in our diary study are also in agreement with the findings of Yeh (2016) who divided participant’s search session into three stages: pre-viewing, viewing and post-viewing. The author differentiated between two ways 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 motivation for active search (passing time, search out of boredom) and motivation of passive search (recommendations by peers, information encountering).

Previous studies in casual leisure information behaviors (Elsweiler, Wilson, & Lunn, 2011) argued that information needs are absent in casual leisure scenarios and the success of these searches is not related to finding relevant information. Our data provided examples for both cases. Some of the casual search sessions initiated with precise information needs while others were simply triggered by hedonic motivations.
6. Discussion

The main contribution of this study is to investigate the use of relevance criteria when searching for videos in leisure context. Previous studies in the area of relevance criteria focused mainly on academic or work related contexts and the information judged was mainly in text format. Thus, this study attempted to fill in the gap of how people assess videos to watch for leisure purposes. 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 limited number of studies. Another challenge in comparing relevance criteria studies is that various methodologies have been applied in relevance criteria literature (Maglaughlin and Sonnenwald, 2002; Savolainen and Kari, 2006).

Our study revealed that many relevance criteria mentioned in the academic or work related contexts can be used in leisure contexts too. As was mentioned previously, the coding scheme used in this study 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 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 we also found Topicality an essential criterion. And we were able to expand the scope of Topicality. The definition of Topicality in our study exceed the limited definition of the match between participant’s query and the retrieved videos to include responses where the video match the participant’s interest as well.

Within the criteria related 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 criterion in making relevance judgment decisions. As video retrieval is different than text and it is hard to identify the author of a video, we found author code not suitable in our 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 where those related to the participants’ beliefs or preferences. The increase of importance of this group of criteria are mainly because of the change in the studies’ contexts. Moving from academic or work related context to leisure context has affected the importance of those criteria that used to play a marginal role in the participant’s relevance judgment decision in the former context. For example, criterion such as Affectiveness played an essential role in relevance judgment in leisure context while it had low frequencies of use in previous literature such as Barry and Schamber (1998).

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.

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 other” criterion in their study, however, Recommended Video in our study has a more broad meaning. In addition to recommendation provided by friends, participants also mentioned recommendations by YouTube, web pages, advertisements in public places and other social media sites (Twitter, Instagram).

There were a group of relevance criteria from previous studies in academic context which did not appear in our data. Examples of these criteria include: Affiliation “relating a document to a particular organization, event, or initiative” (Fulton and Vondracek, 2009), Authority which is a judgement based on references or citations and literature "whether the video has related literature". These criteria tend to be more specific to 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 amount of mentions) varying among different contexts. New criteria emerged which tend to be more related to leisure context.

7. 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 in keep filling out the diary, thus the data
collected from each participant did not necessarily capture 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 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 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 participants say in their diaries. Participants' ability to express their relevance criteria applied in the relevance judgment process varies from one to another and the information provided by the participants were on varied levels of details. In case of brief diaries, the researcher’s 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.

8. Conclusion
This study was undertaken to investigate relevance criteria when searching videos in leisure context. Previous relevance criteria studies were mainly focus on text retrieval in academic or work-related context, thus, a study which investigate video relevance criteria in leisure context was needed. To address this goal a naturalistic diary study was applied and 30 participants filled out the diaries. The analysis revealed 28 video relevance criteria in 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 judgments. As future work, we intend to identify how relevance criteria used during different stages of the search process. This investigation might provide insight of the dynamic aspects of the relevance judgment process.

Acknowledgment
This work was supported by King Saud University, Saudi Arabia. The authors wish to thank the participants who participate in the study, as well as the anonymous reviewers for their constructive comments.

References


<table>
<thead>
<tr>
<th>Date</th>
<th>Time (morning, afternoon, evening)</th>
<th>Location e.g. home, work</th>
<th>Device used e.g. desktop, tablet, mobile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**What are you looking for?** e.g. “muffin recipe”

**Why you start this search?** e.g. kill the time, change my mood, I want a new dish for the dinner

For **three** videos you choose to watch, please write the title of the videos and the reasons that make you select these videos

<table>
<thead>
<tr>
<th>Video # 1 Title</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you stopped watching this video and did not complete it until the end, please give the reasons for that.

<table>
<thead>
<tr>
<th>Video # 2 Title</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you stopped watching this video and did not complete it until the end, please give the reasons for that.

<table>
<thead>
<tr>
<th>Video # 3 Title</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you stopped watching this video and did not complete it until the end, please give the reasons for that.
Table 1
Part of the merged coding scheme from previous literature

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth/Scope/Specificity</td>
<td>Barry &amp; Schamber</td>
<td>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</td>
</tr>
<tr>
<td>Specificity</td>
<td>Savolainen &amp; Kari</td>
<td>The extent to which information is focused enough to match the needs of the user</td>
</tr>
<tr>
<td>Scope</td>
<td>Balatsoukas &amp; Ruthven</td>
<td>For judgments about the depth, scope, completeness, or level of specificity of information.</td>
</tr>
<tr>
<td>Depth</td>
<td>Xie &amp; Benoit</td>
<td>The extent to which detailed information is provided by the document.</td>
</tr>
<tr>
<td>Scope</td>
<td>Xie &amp; Benoit</td>
<td>The extent to which information is covered.</td>
</tr>
<tr>
<td>Specificity</td>
<td>Xie &amp; Benoit</td>
<td>The extent to which information covered by the document is focused to match the user needs.</td>
</tr>
<tr>
<td>Variety</td>
<td>Savolainen &amp; Kari</td>
<td>The extent to which the source provides a sufficient variety of information.</td>
</tr>
<tr>
<td>Coverage(Depth, Scope, Specificity, Variety)</td>
<td>This study</td>
<td>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</td>
</tr>
<tr>
<td>Category</td>
<td>Criteria</td>
<td>Mentions</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Criteria related to information content of the video</td>
<td>Coverage</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Topicality</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>Recency</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Genre</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>People in the Video</td>
<td>50</td>
</tr>
<tr>
<td>Criteria related to participant's previous experience and background</td>
<td>Background/ experience</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Novelty</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>54</td>
</tr>
<tr>
<td>Criteria related to the participant's beliefs and preferences or situation</td>
<td>Affectiveness</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Serendipity/Curiosity</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Habit</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Time Constraint</td>
<td>5</td>
</tr>
<tr>
<td>Criteria related to the quality aspects of the video or the source providing the video</td>
<td>Quality of Source</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Content Quality</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Technical Quality</td>
<td>14</td>
</tr>
<tr>
<td>Criteria related to audio/visual features of the video</td>
<td>Cinematography</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Visual Appeal</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Sound/Voice</td>
<td>20</td>
</tr>
<tr>
<td>Criteria related to the accessibility of the video</td>
<td>Cost</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Language</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Version</td>
<td>3</td>
</tr>
<tr>
<td>Criteria related to other information within the environment</td>
<td>Availability</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Verification</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Unusualness</td>
<td>7</td>
</tr>
<tr>
<td>Criteria related to other people's opinions or YouTube's recommendations</td>
<td>Rank Order</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Popularity</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Recommended Video</td>
<td>72</td>
</tr>
</tbody>
</table>
Table 3
Participants’ topics for leisure searches

<table>
<thead>
<tr>
<th>Topics</th>
<th>Examples</th>
<th>Count of mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>“Adele hello” “pop music”</td>
<td>36</td>
</tr>
<tr>
<td>Beauty</td>
<td>“Hairstyle tutorial” “how to take care of curly Hair”</td>
<td>18</td>
</tr>
<tr>
<td>Tv or YouTube shows</td>
<td>“best of the voice 2015” “The Ellen Show”</td>
<td>14</td>
</tr>
<tr>
<td>Movie trailers</td>
<td>“I know what you did last summer video” “movie snippets”</td>
<td>14</td>
</tr>
<tr>
<td>Celebrities</td>
<td>“Antony Robbins video”</td>
<td>7</td>
</tr>
<tr>
<td>Video games</td>
<td>“looking for a new game to play”</td>
<td>7</td>
</tr>
<tr>
<td>Drama series</td>
<td>“tv series”</td>
<td>7</td>
</tr>
<tr>
<td>Comedy clips</td>
<td>“funny videos”</td>
<td>6</td>
</tr>
<tr>
<td>Cooking</td>
<td>“Sweet potato recipe”</td>
<td>6</td>
</tr>
<tr>
<td>Animals and plants</td>
<td>“strange animal partnerships”</td>
<td>4</td>
</tr>
<tr>
<td>Decoration</td>
<td>“I was, in this case, in search of decoration ideas”</td>
<td>3</td>
</tr>
<tr>
<td>Sports</td>
<td>“to look at sports and people”</td>
<td>3</td>
</tr>
<tr>
<td>Travel and tourism</td>
<td>“Okun island”</td>
<td>3</td>
</tr>
<tr>
<td>Motivational speeches</td>
<td>“motivation video”</td>
<td>3</td>
</tr>
<tr>
<td>other</td>
<td>“Videos about US presidential race” “magic tricks”</td>
<td>13</td>
</tr>
<tr>
<td>Motivations</td>
<td>Examples</td>
<td>Count of mentions</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Pass Time</td>
<td>&quot;Getting something to watch during waiting time&quot;</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>&quot;Kill the time&quot;</td>
<td></td>
</tr>
<tr>
<td>Change Mood</td>
<td>&quot;change my mood&quot;</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>&quot;I was feeling bored and I wanted something to cheer me up&quot;</td>
<td></td>
</tr>
<tr>
<td>Relaxation and Refreshment</td>
<td>&quot;wanted to relax after a long day and before going to evening work&quot;</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>&quot;Take a break and refresh my mind while working on a website&quot;</td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>&quot;I want to learn how to make a natural look for everyday&quot;</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>&quot;wanted to learn a new dish for dinner&quot;</td>
<td></td>
</tr>
<tr>
<td>Background or Distraction</td>
<td>&quot;my usual background when I am doing body exercise&quot;</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>&quot;to block out building noises in my house which is having a loft extension&quot;</td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td>&quot;a friend recommended the band to me so I checked them out&quot;</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>&quot;I would like to watch the trailer of the latest movies, because posters of some of new movies look quite interesting&quot;</td>
<td></td>
</tr>
<tr>
<td>Entertainment</td>
<td>&quot;for some laughs&quot;</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>&quot;I went on YouTube purely seeking entertainment.&quot;</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>&quot;our teacher showed it to us and I want to show it to my sisters&quot;</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>&quot;having trouble sleeping&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;my daily habit of checking my subscriptions&quot;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;curious to hear the song that everybody is singing at the university&quot;</td>
<td></td>
</tr>
</tbody>
</table>