

A Literature Review of Personalisation, Search, and Navigation Methods in Open e-Learning

Dhuha Al-Shaikhli and Simon Courtenage

Abstract—Learning engagement has been a significant field of research during the past decades, especially in terms of e-learning. Personalisation, search, and navigational methods have been reported as potential approaches to supporting learners on e-learning platforms. In this review, we discuss previous research on the above-mentioned fields and recent results that e-learner providers can benefit from.

Index Terms—E-learning, learning environment, navigation, personalisation.

I. INTRODUCTION

In the past, navigation mainly depended on the hierarchical structure of creating directories or metaphoric folders, which represented information retrieval in the digital world [1]. Some researchers have categorised the navigation sequence process into two types: forward search and impasses [2]. In the forward search process, the user follows symbols that either literally match or are similar/relevant to the context of the user's goal or represent a context that indicates the user's goal (i.e. where they are aiming to reach in their search process). In the forward search process, the user moves to the next page by choosing the screen object that is relevant to their goal. On the other hand, an impasses search is performed when the user reverses their navigation or backtracks, as they reach a page where they cannot find the closest similarity (in terms of links or tabs) to their search goal.

Many decades ago, users created their own hierarchical folders and used them to store their files. The information retrieval journey goes through the folders' hierarchy in order to reach the destination one and locate the required file [3]. This technique was criticised due to its slow information retrieval process, as it depended on user memory and on the names of files and folders as defined by individual users, rather than any standardised folder naming system. It would 'hide' the information contained within the folders in the hierarchy. In addition, file categorisation became challenging with a larger number of files [1].

Recent research shows that users are not efficient in organising their files, and the structures they follow for information retrieval via the navigation approach are shallow. However, semantic and structured folders and the limited depth of the folders (i.e. between three and four folders) improve retrieval time and the success of files [3]. Similar findings show that a deep structure for online learning

material reduces learner engagement or means that the files are not viewed at all [4]. Similarly, in another work by [5], some navigation patterns show that learners decrease in engagement or in the number of visits as they move downward from the home page to a specific lesson on the course. In other words, the more distant a learner is from the home page, the less engaged they are with the deeper content. This finding corresponds positively with [2]'s study, which noted that the depth of the navigation route on the web decreases the possibility of success for the user reaching their goal. A graphical dynamic and interactive map showing the footprint of the user through the pages they have visited could support user engagement; it could also be used as a shortcut to a particular page on the web or as a reminder of what the user has viewed recently.

Moreover, learning materials should be pre-planned and allocated, as website users do not engage with content-concentrated web pages [4]. In a study carried out by [5], the maximum number of contents viewed was 400 when the number of users was 2926, and the contents were viewed by 1–5% of learners. More interestingly, no single contents were viewed by all learners. Reference [5]'s results show that most of the course contents were accessed by students. This finding is considered important in this field, as previous research showed that fewer students accessed their e-learning courses. Reference [5] reasons that the increase in the number of students accessing course contents was due to the course websites being structured according to various approaches—for example, a chronological order for course instructions or based on types of learning material. Therefore, as the contents were not organised linearly, learners were able to explore a wider range as they searched for particular content. Another major finding in his research was the variety of learners' access and navigation techniques, which indicates individual preferences and learning styles. Recently, personalisation and individual preferences have been considered when developing online learning material in order to cater to a wide range of learning approaches [4].

II. NAVIGATION

Navigation began by depending mainly on the hierarchical structure of creating directories or metaphoric folders, which represented information retrieval in the digital world [1]. Some researchers have categorised the navigation sequence process into two types: forward search and impasses [2]. In the forward search process, the user follows symbols that either literally match or are similar/relevant to the context of the user's goal or represent a context that indicates the user's goal (i.e. where they are aiming to reach in their search

process). In the forward search process, the user moves to the next page by choosing the screen object that is relevant to their goal. On the other hand, an impasses search is performed when the user reverses their navigation or backtracks, as they reach a page where they cannot find the closest similarity (in terms of links or tabs) to their search goal.

Many decades ago, users created their own hierarchical folders and used them to store their files. The information retrieval journey goes through the folders' hierarchy in order to reach the destination one and locate the required file [3]. This technique was criticised due to its slow information retrieval process, as it depended on user memory and on the names of files and folders as defined by individual users rather than any standardised folder naming system. It would 'hide' the information contained within the folders in the hierarchy. In addition, file categorisation becomes challenging with a larger number of files [1].

Recent research shows that users are not efficient in organising their files, and the structures they follow for information retrieval via the navigation approach are shallow. However, semantic and structured folders and the limited depth of the folders (i.e. between three and four folders) improve retrieval time and the success of files [3]. Similar findings show that a deep structure for online learning material reduces learner engagement or means that the files are not viewed at all [4]. Similarly, in work by [5], some navigation patterns show that learners decrease in engagement or in the number of visits as they move downward from the home page to a specific lesson on the course. In other words, the more distant a learner is from the home page, the less engaged they are with deeper content. This finding corresponds positively with [2]'s study, which noted that the depth of the navigation route on the web decreases the possibility of success for the user reaching their goal. A graphical dynamic and interactive map showing the footprint of the user through the pages they have visited could support user engagement; it could also be used as a shortcut to a particular page on the web or as a reminder of what the user has viewed recently.

Moreover, learning materials should be pre-planned and allocated, as website users do not engage with content-concentrated web pages [4]. In a study carried out by [5], the maximum number of contents viewed was 400 when the number of users was 2926, and the contents were viewed by 1–5% of learners. More interestingly, no single contents were viewed by all learners. Reference [5]'s results show that most of the course contents were accessed by students. This finding is considered important in this field, as previous research showed that fewer students accessed their e-learning courses. Reference [5] reasons that the increase in the number of students accessing the course was due to the course websites being structured according to various approaches—for example, a chronological order for course instructions or based on types of learning material. Therefore, as the contents were not organised linearly, learners were able to explore a wider range as they searched for particular content. Another major finding in his research was the variety of learners' access and navigation techniques, which indicates individual preferences and learning styles. Recently,

personalisation and individual preferences have been considered when developing online learning material in order to cater to a wide range of learning approaches [4].

III. SEARCH

Eventually, search techniques were invented and these provided easier and more efficient information retrieval. With search facilities, users no longer needed to remember storage folders and how to reach them [1]. Interestingly, some research studies showed that users preferred navigation to search facilities [6]. This could be because these studies were performed in the early days of the invention of search facilities, when they were not as improved, or it could be due to human nature (people are more secure when they have overall control; for example, in navigation, the user can reach the basic root of their files). Another reason could be that, in searching, the user needs to recall information, whereas navigation requires recognition. Recognition tasks are usually simpler and need less cognitive effort [1].

IV. PERSONALISATION

A. Introduction

A MOOC, which is a good example of open e-learning, is meant to serve a significant number of learners from a wide range of cultures and backgrounds; therefore, it is essential that it is equipped with personalisation attributes in order to meet a wide range of individual needs and preferences to some level for any field [7] and particularly for e-learning [8]. Personalisation in web-based interactive environments has been considered a key element in enabling effective and efficient user engagement [9]. This has led Amazon, Google and Bing to start employing various forms of user modelling to implement and provide personalisation attributes for their users. Nowadays, Google users always find that ads of their recently viewed products appear on any website they visit, which is called content filtering personalisation [10]. There are various types of automatic recommendation systems that are fed by user activities, and these personalise each user's web experience with specific recommendations based on their recently viewed or liked content or content they've interacted with.

In terms of personalising an e-learning environment, as e-learning material is accessed by learners with various learning styles and cognitive attitudes, research emphasises the importance of implementing personalisation and user adaptation to provide learners with a positive learning experience [4]. In e-learning, personalisation applies to the adaptive environment for the learner [7]—i.e. adaptive collaboration, adaptive content delivery and adaptive resources [10]. An essential attribute for personalisation in web interactive systems such as a MOOC is that the system must automatically assess and update user profiles, as their preferences may change and hence their customised interface and content presentation would need to be adjusted as well [9].

B. Previous Research on Personalisation

The internet has been used for decades as a medium of sharing information; however, since it has been used by a significant range of users with diverse perspectives, online interactive systems (especially the educational ones) cannot globalise their content delivery methods into a single approach; they must cater to the various perceptions and learning and cognition styles [9]. Such homogenisation has been one of the main reasons why learner active engagement has been limited in e-learning environments. Therefore, researchers have recently proposed adaptive systems that adjust based on specific user functioning and attributes in order to increase task completion and active engagement in web-based interactive systems [9]. For example, [9] applied a psychometric-based assessment to tackle human navigation behaviour in a way that reflects a user's cognitive style for e-learning. The same concept has been discussed by [11], who stated that the navigation process involves cognition. On the other hand, [8] followed a different approach in the same field by implementing a recommender system for the user based on their own experiences in order to personalise their sessions by providing adaptive navigation support. The system caters to each user's preferences, and reduces the high noise of available content on the web. As a result, it improves user's engagement with the environment. In general, the recommender system is a program that utilises information from the user's data mining in real time and, based on their activities and explicit information, provides them with action recommendations [12]. Although [8] share with [9] and [11] the importance of considering the learner's cognitive style, their research took a different path by accumulating users' explicit information. Their research involved feedback from learners and instructors in the feasibility stage in order to address unseen obstacles and gaps in the system. Such feedback provided a clearer picture on how to cater to learners' needs while they interact with the e-learning environment [8]. The researchers' purpose was an adaptive navigation support system that personalises learners' experiences, which, in turn, supports their active engagement.

Reference [9] aimed to identify any user-shared attributes in terms of navigation and search behaviour when users are exploring an e-learning environment. The research tackled a wide range of learners' perceptual, functional and cognitive styles available in an e-learning environment in order to create diverse acceptance and active engagements by enabling users to complete particular tasks more efficiently and effectively. Based on this, [9] discuss adaptive interactive system functionality with a static user interface. Their proposed system was implemented in user-centred design (UCD) methodologies, which position the user at the centre of the development process. This is the same approach followed by [8] for their implementation of a personalised e-learning environment. These UCD methods include design and weigh adaptive mechanisms that cater to a user's cognitive style based on the fact that a learner utilises their cognition to process learning material [9].

A user-adaptive system needs to be able to identify the user's activities that exhibit their type of interaction, as well as the preferences that could be analysed at a later stage to define other specific individual attributes and cognition

styles. For example, a user's navigation behaviour could be either holistic or analytic. In the first approach, the user references the main content using a linear approach or a simple-to-complex approach [8] in order to absorb meaning. An analytic style is when the user relies mainly on absorbing the meaning of the available content first, before navigating to the next topic based on the result of their analysis. These two attributes may be shared to form a unique group of users. Therefore, the system needs to adapt based on the group's features in order to match their needs, increase efficient interaction and provide them with a positive learning experience [9]. One of the methodologies used to build user models that track and log user navigation patterns is to log user interaction using hyperlinks. Each hyperlink has related content attached to it; hence, as a user interacts with these links, they indicate the user's cognitive and navigational style and whether they have a linear approach to navigation. Interestingly, researchers found that users who share the same navigation pattern do not necessarily share the same cognitive style. This could be either due to the preliminary stage of the available cognitive style identification systems, or it could be due to the uncertain and complex human approach to processing information via other facts that cannot be tracked, at least not via the web [9].

In [9]'s research, the user model is where the adaptive system utilises user information in order to adjust and personalise to suit each user based on their stored static and dynamic information (which has been logged explicitly or implicitly). It does so in order to meet their needs, preferences and analysed perceptions, which is the same approach followed by [10]. Usually, the method of user explicit information—or as, [10] defined it, collaborative learner modelling—can be applied through an online questionnaire, where users provide direct answers. For example, several user cognition styles can be accumulated using the Index of Learning Styles (ILS) questionnaire by Felder [3]. On the other hand, implicit information—or automatic learner modelling [10]—can be gathered by recording user downloads and content ratings, which provide insight on the specific material the user is interested in. Simple implicit information gathering is based on user online behaviour and activities. For example, [9] state that implicit information may be gathered through the history of the user's navigation, their saved searches, viewed links, and time spent on specific web pages (which reflect the user's interest in a particular field). Some researchers favour implicitly gathered information, as it is collected in the background and does not disturb the user's natural functioning. In addition, it is an automated process from the start till it ends on the user's model database [9]. However, explicit information gathering is still heavily relied upon due to the direct information it provides about users [10]. In both methods, in order to transform the accumulated data into useful data, web mining [13] techniques are required to produce a clear picture of the user patterns in specific interactive web environments. After this, field experts can plan how to support users to ensure a positive experience [10].

The most common implicit information gathering method is tracking the user's browsing history and the length of time each URL has been visited (which exhibits clear user

interest). The obstacle for collecting browsing history data is that an initial setup is required; this can be on the user's machine and needs user authentication (and thus can be ignored by most users [9]). Another method of implicitly gathering user data is using agents such as plug-ins. These agents collect actions performed by the user, such as downloads and bookmarks. The browsing agent may guide the user to their preferred resources based on the information it collects from the user's activities. However, browsing agents require the user to install specialised software in order to function, which can be ignored or rejected by a significant number of users. Another method of user modelling through implicit information gathering is when using the frequently used topic or link of content without processing them and adapting them on them to be displayed on the user interface.

Some researchers tend to use explicit user information gathering methods, as they are direct and clear. In addition, explicit information needs less analysis and is less complex—for example, when a user states that they would like to study a HTML programming course, it is more useful than if we viewed gathered information on the user's model database that exhibited 15 minutes spent on an HTML programming document. In the latter case, it is not clear if the user engaged with the viewed document or skimmed it. A combination of explicitly and implicitly gathered information may provide a stronger user model—for example, asking a user through a web questionnaire what programming language does they like and also tracking their practice and viewing activities in this field to verify their preference. User-guided modelling methodologies aim to gather information directly from the user, which could be via an online questionnaire or registration form. Yet there are disadvantages of user-guided or explicit information gathering methodologies: the user may not be willing to provide accurate information due to a lack of trust or time. In addition, the process of filling out a form while or prior to engaging with an interactive system may be disruptive and hence affect the accuracy or completion of the provided information [9].

Reference [9] focuses on mainly extracting and analysing the cognitive aspect of user navigation and the activities of viewing, downloading, etc. The researchers applied various user information gathering methods in order to have a better understanding of user perceptions when they explore an e-learning environment. However, the research would have been significantly complemented if it involved the crucial aspects of content layout, design and structure, time allocation for tasks, and task repetition to retain information, and analysed their impact on learner cognition, encoding of information and acceptance. Reference [11] discusses how a learner's cognition is significantly influenced by content layout, position and structure, and how these aspects work with the learner's long-term memory, deep learning process, and information retrieval and navigation processes. Reference [9]'s adaptive system would reach another level of understanding and improving learners' cognition styles if it involved [11]'s research aspects.

Researchers and specialists in the field of user navigation and personalisation have developed various techniques and systems that support the user's approach to the content they are looking for. For example, a personal preferences history accumulator that interacts with hybrid search navigation technology is one of the most widely utilised techniques, including on major websites such as Amazon and Google. Based on the user's personal preferences and an activity history, resources and frequently visited web pages are displayed as suggested hypertext links at the user's front page or on the build in a search box [14], like on Amazon. A search box that autocorrects [15]–[17], like on Google, allows the user to retrieve what they have recently searched for as a personalised search [14], and the results are followed by displaying the most commonly searched for term [16]. Alternately, the user may search for highly rated discussion topics or articles [18].

Normally, the search bar employs correlations with the search terms to amend misspelled words in the query [15]. The correlation shows the recurrent search terms that have historically appeared together in a specific query, and have been saved in the correlation database while utilising relevant terms lists. The correlation database is generated recurrently from recent entries in a query register. Hence, it presents the user's recent preferences [15]. In the same field, a search shortcut method has been implemented in the search box to decrease the time spent on information retrieval [17]. Researchers like [17] have implemented a shortcut facility that saves the recently visited information and redirects the user to them via a shortcut without going through the same sequence they use every time [17]. It is essential that the search facility is able to retrieve different types of files from across different databases.

Another technology that has been implemented in the search and navigation field is the filtering technique. Filtering information narrows retrieved information based on date, type, key words, etc. Usually, the search facility needs to have an incremental search attribute. An incremental search allows the search process to start as soon as the user enters the first letters of the key word or sentence. Thus, results are listed based on the words entered by the user and the best match is displayed at the top; thus, it provides faster results for the user. The incremental search places recognition over recall, which makes it easier for the user to find information [1].

In e-learning, navigation is mainly implemented and structured based on resources and interactive learning spaces, such as collaboration platforms, interactive quizzes, etc. Navigation should be well planned and focus on the learning material in order to support the reader's visual concentration. High weight should be given to the material layout and the articles' lengths and structures in order to allow learners to construct their knowledge without getting distracted or confused [19]. The layout and navigation of the material should follow an approach that allows the learner to anticipate the topic or knowledge they are going to read next [19]. The hypertext organiser is one suggested approach by [19] and has been used by Wikipedia, as each article contains a hyperlinked table of contents. The table of contents allows a larger amount of web pages within an article and thus

includes more relevant material about the topic. The user can then navigate using a structured approach; this approach is known as intra-article navigation. The text's devices aid the reader's ability to develop a mental structure of the article's material [19]. Similarly, other researchers found that e-learners engage more and have a deeper level of processing with textual information when it is positioned on the left-hand side of the web page; this makes sense, as reading is from left to right [11]. In the same field, researchers discussed the importance of implementing structured text allocation, length and navigation in order to support the reader's visual engagement and to enable them to relate and reflect on the text they are reading, as well as construct knowledge as they read [19]. For example, some researchers found that learners absorbed more knowledge via linear text than hypertext, and better retained information when reading from a shallow to deeper context in an e-learning environment [11].

VI. DATA AND WEB MINING

Data mining is an emerging method that has been significantly used in e-commerce to increase profits and in e-learning to improve teaching and learning [5]. It is a set of tools and techniques that are used to find unpredictable or hidden patterns in users' activities while they interact with an online system [6] pp 1–17. For example, a specific navigation pattern exhibits users' specific access routes to the learning material and how they navigate among that material. This helps inform future improvements on the contents' structure to create a smoother route for user navigation, which, in turn, supports user engagement [4]. Researchers defined the general steps of data mining as the following. The first step is pre-processing, where the obtained data (from the database of the environment that the user interacts with) undergo a number of processes, such as cleaning, attribute selection, transforming attributes, data integration, etc. This is followed by the data mining process, which is the main stage; here, data go under several data mining techniques, such as visualisation, regression, classification, clustering, association rule mining, sequential pattern mining, text mining, etc. The final process is post-processing, where the results of the interpreted data are used to understand the user's pattern; this is followed by decision making [20]. The association rule—which is one of the most heavily used techniques in data mining—is a clustering technique that basically recognises the correlation among learning contents depending on their co-occurrence through sessions. The association rule helps in predicting specific user behaviour, as well as reveals a probable relationship between contents that users interacted with [12].

Web mining is a technique of data mining that gathers the relevant information associated with web page retrieval in order to create a web session for eventual analysis. For example, web mining provides a better prediction of user access and the usage of web pages [13]. This type of information is essential to plan sustained active engagement for MOOC model resources and course content structure in order to complement users' attitudes towards navigation. In addition, web mining provides vital insight into e-learners'

attitudes towards the material; therefore, educators and tutors can plan their pedagogical strategies accordingly [13]. However, it is crucial to determine what parameters are actually affecting learners' attitudes when they interact with the e-learning environment and whether these parameters are approachable via an online environment. Despite this, data mining still provides significant information that can help improve the learning environment. In terms of learner engagement with the e-learning environment, [5] and Hwang and Wang defined six behavioural patterns among learners: persistent, according to recurrent visits and time frames; prior; prior plus; midterm; prior-posterior; posterior; and posterior plus. These patterns present the phases of learning, when the learner shows interactive engagement. For example, prior learners refer to those students who are actively engaged in the early phase of the course, whereas prior plus learners exhibit longer periods of active engagement than prior learners [5]. These results help to at least define the active learners (prior and prior plus), and we can now explore their activities and investigate when they start to drop off the course.

Researchers have proposed various techniques for implementing a better web log [21]. Mainly, logged data analysis is concerned with three main parameters: the action taken, who performs it, and when [5]. A log file is any accessed or downloaded content on the web, which will be automatically logged as a text file with an identical format on the background database. The pre-processing phase of web log data is one of the essential stages of web mining. The phases in web mining are: data collection, data pre-processing, and pattern discovery and pattern analysis. Pre-processing web log data (which are users' files index interaction with the web page) is about extracting the sequence of recurrent navigation patterns; this provides tangible predictions for developers and educators to consider in terms of content structure [21]. Learners' activities that can be counted through data mining are mainly logins, page views, collaboration on forums, and content downloads [5]. This phase of data pre-processing has received wide attention in the research and literature, as it is a key phase in determining the accuracy of the pattern that will be produced from the collected data.

Another approach of data mining and learners' log files was implemented by [22] based on a time correlation with learners' performed tasks. This approach starts by identifying each student with a serial number and logging this number in the database. Then, each file on the database row starts with the learner's action of accessing the system, and ends with the closing of the application or navigating away from the environment. Between these two actions, there is a set of middle actions performed by the learners while they are interacting with the e-environment. Usually, these actions are filtered from some reasonably irrelevant actions, such as learners who accessed the environment and performed less than three actions or who viewed an article for more than an hour. After the cleaning process, the collected data go through the analysis phase, where patterns are distinguished and matched [22].

Measuring a learner's pace is another topic explored by [22], who accumulated users' data and traced them in the

e-learning environment to analyse them for future improvements in the system. The research study chose to tackle a time-related variable, where it could exhibit learner engagement with visual content, quizzes or other forms inside an e-learning environment. Pace (speed, rate) has been defined by researchers as “the number of activities completed, divided by total time on task – tend to be more cognitive than behavioral” [22]. As a term, ‘pace’ has been interpreted as having two meanings: ‘pace of learning’, which is “the time is taken to complete a task” and ‘pace of action’, which is the number of actions per time unit. Both pace of action and pace of learning are interrelated in one way or another. In addition, the nature of the tasks affects both concepts of pace. Therefore, previous research has characterised pace for students based on the whole learning period.

Interesting results were presented by [22]’s study, which stated that a learner’s pace (for both meanings) varies significantly depending on the time of day and night. In addition, learners show higher performance at the end of sessions, as they tend to increase their focus in order to complete their tasks. Another finding that appears in [22]’s research is that the learning mode varies for each student, such as memories, practice, etc. However, the correlation between tasks and pace remains scattered among students, and there was no significant consistency in any relation between tasks and pace in general in the e-learning environment. Reference [22] reasons that the low consistency between tasks and pace is due to the need to further investigate this relationship (which might show a stronger bond to learner’s pace consistency). In addition, other components—such as individual perception and previous knowledge—could highly affect task performance and engagement with online content [22].

In another research study on data mining for e-learners, Nachmias examined learners’ types of persistence and divided this into volume of activity and direction of activity. Following data analysis, a number of patterns linking these two independent axes were found [5]. Accordingly, Nachmias defined five types of learners based on their online active engagement: low extent users, late users, online quitters, accelerating users, and decelerating users (which is the largest group, as they form 33% of learners). The decelerating users exhibit an overall decrease in their online active engagement during the course. Nachmias suggested that the number of e-learning users tends to decrease during the online course, and that this needs to be further investigated and understood. Most of the learners are online quitters and in the decreasing group; these tend to access the course content and interact with it at the beginning of the course. They form 46% in Nachmias’s research study and 42% in Hwang and Wang’s research study. For prior learners and prior plus learners (both of whom Nachmias referred to in his work), the percentages are close to each other. The following largest group is accelerating users (as they formed 29% in Nachmias’s research), who are learners who access course content and actively interact with the material in the last phase of the course (i.e. towards assignment submission deadlines and exams). Nachmias’s research results were similar to his peer researchers, Hwang and Wang, and he referred to their work in his study [5].

VII. CLUSTERING METHODS

Various methods of clustering data have been suggested. The hierarchical agglomerative clustering technique is a bottom-up approach where each sample is treated separately and is merged with a pair sample (based on the maximum value of their recurrence user interaction with web pages), until all clusters are merged together into one single group [21]. The hierarchical agglomerative technique groups patterns based on the sequence of occurrence of web pages. Once the system has grouped all the possible pages based on visits as well as sequence, it is able to exhibit a successful user-supporting web content structure [21]. Another well-known algorithm for user clustering is K-mean, which was applied in [10]’s research. This method is concerned with clustering users based on their web sessions. Researchers have focused on the K-mean clustering algorithm and have proposed various improvements, such as clustering web sessions based on increased similarity. Others have suggested clustering web sessions based on the usage of web pages. Furthermore, the trajectory clustering technique has been suggested to cluster users based on the dissimilarity of their web sessions [21].

Researchers are implementing programmed data mining in their user models in order to cater to the large amount of user information [9]. For example, in her literature review, [9] refers to a researcher who applied the clustering process on users’ sessions to locate users who shared the same activities; the researcher was then able to create a particular structure for these users, such as a collaborative group [12]. This is the same approach followed by [10]. Reference [10] has extended this approach by modelling web content so that it is based on user access and sessions, and provides standardised educational content as part of a recommendation system. However, neither of these studies has considered content layout and text positioning as significant aspects that could impact learner comprehension and the level of processing (as discussed earlier for [19]’s and [11]’s work).

Reference [9] discussed how some researchers cluster their user models based on navigation behaviour [9]. The method behind this clustering process is to separate a set of activities or type of navigation behaviour into segments of random users, then constantly match these segments with similar ones. A segment matches with the most similar segment, and then keeps searching until it finds the best similarity in that segment and it classifies users based on their matching segments. This is similar to [10]’s approach, as the researchers processed their users into two levels depending on similarities and dissimilarities among favoured learning content. Some users were classified into more than one set, which is identified as a fuzzy clustering technique. This technique feeds the adaptation system with clearer navigation predictions for users with similar profiles [21]. A data analysis of navigation patterns may identify hidden relations between resources and contents; for example, if a considerable number of users always navigate from X to Y, then this information helps in restructuring X and Y to create a more beneficial relationship. In addition, Y could be recommended by the system for users from other segments that might benefit from it as well. Hence, an adaptive system that tackles the cognitive aspect may suggest different

content presentations and structures based on the majority of users' preferences or behaviour clustering [9].

Similar to [9], other researchers have started to tackle and employ users' cognitive styles as major criteria for clustering and analysing their user models on interactive web systems in general. Many researchers have acknowledged that cognitive styles do interfere with user performance on an online interactive system [9]. Unlike learning styles, cognitive styles exhibit the personalised strengths and preferences of individuals regarding how they perceive, process, and organise information. This affects the user's capability of accessing and processing information. In contrast, a learning style has been defined by Felder as "The ways in which an individual characteristically acquires, retains and retrieves information" [23].

Hence, [9] proposed modelling user's cognitive styles based on group of web usage mining methods on user navigation pattern. Researchers could analyse cognitive styles using users' behavioural patterns, which could be obtained through navigation pattern analysis or self-reporting through an online questionnaire, combined with the study of other users' intellectual information. Based on the literature review made by [9], the dimensions of cognitive style have been defined as: "field dependent/independent, impulsive/reflective, conceptual/inferential, thematic/relational, [and] analytic/global." There are particular cognitive styles that are more measurable via web content and tools [9]. Researchers have focused on implementing adaptive interactive online systems that utilise personalised cognitive style dimensions, such as Riding's (1991) cognitive style analysis (CSA) system referenced by [9].

In addition, [9] discussed various user modelling mechanisms, where they mainly tackle learners' cognition or learning styles. For example, CSA (verbal/imagery and holistic/analytic) has been utilised by some researchers to define visual and verbal users, and holistic and analytic ones. Another example is Felder's ILS, which has four scales: active vs reflective, sensing vs intuitive, visual vs verbal, and global vs sequential [3]. As discussed by [3], ILS is a self-scoring questionnaire answered by learners; it shows which one of the four scales they are positioned in, and whether they are flexible or strict learners within that particular scale. The ILS does help in recognising and clustering learners in an e-learning environment, if answered completely and correctly. The questionnaire could guide us towards recognising potential course dropouts and how they can be supported. However, ILS has 44 multiple choice questions, which means that there is a chance it will not be completed. In addition, there is a chance of inaccuracy among the chosen answers, as learners may repeat a certain choice in order to complete the questionnaire faster.

VIII. CONCLUSION

Methods to support learning engagement are more efficient when combined. For example, a user benefits more from a personalised auto-completion search bar that retrieves results based on the user's previous experience (derived from data mining and clustering methods). The new series of

personalisation methods consists of more adaptive systems, which respond to the user in real time and interact with the learner's psychological patterns. Thus, they promise to provide more effective support to users [24].

REFERENCES

- [1] O. Bergman, R. Beyth-Marom, R. Nachmias, N. Gradovitch, and S. Whittaker, "Improved search engines and navigation preference in personal information management," *ACM Transactions on Information Systems*, vol. 26, no. 4, pp. 1–24, 2008.
- [2] M. Kitajima, M. H. Blackmon, and P. G. Polson, "A comprehension-based model of Web navigation and its application to Web usability analysis," in *People and Computer XIV – Usability or Else!*, Y. W. Sharon McDonald, Ed. London: Springer, 2000, pp. 357–374.
- [3] O. Bergman, S. Whittaker, M. Sanderson, R. Nachmias, and A. Ramamoorthy, "The effect of folder structure on personal file navigation," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2426–2441, 2010.
- [4] R. Nachmias, and L. Segev, "Students' use of content in Web-supported academic courses," *The Internet and Higher Education*, vol. 6, no. 2, pp. 145–157, 2003.
- [5] A. Hershkovitz and R. Nachmias, "Online persistence in higher education web-supported courses," *The Internet and Higher Education*, vol. 14, no. 2, pp. 98–106, 2011.
- [6] S. Hardof-Jaffe, A. Hershkovitz, H. Abu-Kishk, O. Bergman, and R. Nachmias, *How do Students Organize Personal Information Spaces?*, Spain: ERIC, 2009, pp. 250–258.
- [7] K. R. Permlatha, B. Dharani, and T.V. Geetha, "Dynamic learner profiling and automatic learner classification for adaptive e-learning environment," *Interactive Learning Environments*, pp. 1–22, 2014.
- [8] O. C. Santos, J. G. Boticario, and D. Pérez-Marín, "Extending web-based educational systems with personalised support through User Centred Designed recommendations along the e-learning life cycle," *Science of Computer Programming*, vol. 88, pp. 92–109, 2014.
- [9] M. Belk, E. Papatheocharous, P. Germanakos, and G. Samaras, "Modeling users on the World Wide Web based on cognitive factors, navigation behavior and clustering techniques," *Journal of Systems and Software*, vol. 86, no. 12, pp. 2995–3012, 2013.
- [10] M. Jemni and O. Nasraoui, "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval," *Education Technology and Society*, vol. 12, no. 4, pp. 30–42, 2009.
- [11] T. Dikbas, and A. Altun, "The effect of levels of processing with navigation design types on recall and retention in e-learning environments," *Behaviour and Information Technology*, vol. 33, no. 10, pp. 1039–1047, 2014.
- [12] C. Romero, S. Venturaa, A. Zafraa, and P. de Brab, "Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems," *Computer & Education*, vol. 53, no. 3, pp. 828–840, 2009.
- [13] M. Narvekar and S. S. Banu, "Predicting user's web navigation behavior using the hybrid approach," *Procedia Computer Science*, vol. 45, pp. 3–12, 2015.
- [14] E. Dominowska and R. J. Rango, U.S. Patent 7,810,035 B2, 2010.
- [15] D. E. Bowman, R. E. Ortega, M. L. Hamrick, J. R. Spiegel, and T. R. Khon, U.S. Patent 6,144,958, 2000.
- [16] R. E. Ortega, J. W. Avery, and R. Frederick, U.S. Patent 6,564,213 B1, 2003.
- [17] B. Rakowski, G. Murphy, B. Goodger, P. Kasting, and A. Weber, U.S. Patent 8,438,148 B1, 2013.
- [18] F. Mushtaq, G. Das, and A. M. Habib, U.S. Patent US 8,954,867 B2, 2015.
- [19] E. Cuddihy and J. H. Spyridakis, "The effect of visual design and placement of intra-article navigation schemes on reading comprehension and website user perceptions," *Computers in Human Behavior*, vol. 28, no. 4, pp. 1399–1409, 2012.
- [20] E. Garc á et al., "A collaborative educational association rule mining tool," *The Internet and Higher Education*, vol. 14, no. 2, pp. 77–88, 2011.
- [21] D. S. Anupama and S. D. Gowda, "Clustering of web user sessions to maintain occurrence of sequence in navigation pattern," *Procedia Computer Science*, vol. 58, pp. 558–564, 2015.
- [22] A. Hershkovitz and R. Nachmias, "Consistency of students pace in online learning," in *Proc. International Conference on Educational Data Mining (EDM)*, pp. 71–80, 2009.

- [23] D. El-Hamoudova, "MOOCs motivation and communication in the cyber learning environment," presented at the 3rd World Conference on Educational Technology Researches, pp. 29–34, Turkey, vol. 131, 2013.
- [24] H. M. Truong, "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities," *Computers in Human Behavior*, vol. 55, pp. 1185–1193, 2016.

S. Courtenage is PhD computer science, University College London, London, UK, 1995. His current research interests include the study of the navigational behaviour of online learners. He is a principal lecturer in the Dept. of Computer Science, University of Westminster, London, UK.



Al-Shaikhli is MPhil/PhD candidate in computer science, University of Westminster, London, UK. She got the MSc advanced software engineering, University of Westminster, London, UK, 2009–2010; 2013–present; the BSc computer science, Abu Dhabi University, Abu Dhabi, UAE, 2005–2008. She is a visiting lecturer at the University of Westminster, London, UK.