A Semantic Analysis of the Learner's Disorientation

S. Ait Adda, N. Bousbia and A. Balla

National High School of Computer Science, Algiers, Algeria

Abstract— **Learning systems are dedicated for learning about a particular area organized through non-linear documents. Therefore, it is always useful to recognize the state of knowledge and the navigation behaviour of a learner in order to evaluate customize and adapt the learning process. In this paper, we aim to make a semantic analysis of the learner's navigation during his apprenticeship in hypermedia content. The main reason of this analysis is to identify the browsing behaviour of a learner with the current course. We assume that if the semantic distance between the domain concepts of the documents (or pages) that follow each other in the navigation of a learner is great, then this reflects the unstructured navigation behaviour and interprets that the learner is disoriented. This type of behaviour could be due to the poor organization of the content and the bad structuring of the course. Indeed, this analysis will allow to the tutor, to identify the disoriented learners and help them, and to the course author, to specify the causes to restructure and to deepen the analysis of the existing content and the navigation links between the parts of the course.**

Index Terms—**Conceptual indexing, Disorientation behaviour, Domain ontology, e-Learning, Semantic similarity, Trace.**

I. INTRODUCTION

As learning technology, Learning Systems (LS) are being increasingly used in Higher Education [1] to support students' access to learning material in a flexible way. One of the major reasons of moving from traditional classroom-based learning to offer instruction through the use of Learning Systems is that the latter can present learning material in a non-linear structure [2]. Such non-linearity affords learners greater flexibility in navigating the learning content and allows them to choose freely their own paths through it to meet their learning goals [3]. Nevertheless, non-linearity can be also the source of several problems.

The non-linear structure of the course may increase complexity, as learners need to make extra efforts to - understand which content is relevant to their instructional objectives, and to develop a clear idea of the relations among the pages/links and their position in the environment [4]. So, learners may feel confused to interact with the Web-based learning systems. One of the major confusion in which we are interested in this paper, was disorientation experience.

Disorientation is the cognitive effect produced with the reader who does not make any links between his initial browsing project and the zones of information that he has read [5]. Originally, disorientation is often due to cognitive overload resulted from the incapacity of the learner to retain the main part of the information related to a browsed node, or its difficulties to quickly identify the nature of links that he has activated [6]. The learners must identify, orient themselves and navigate through an information space in constantly changing. This does not seem to tangle with slow network growth. As consequence of disorientation is that the learners pay less attention to their learning tasks and can miss at least some of the relevant content in the system, which may hinder their learning performance [7], [8].

In front of these obstacles (cognitive overload and disorientation) designers propose supports of navigation assistant [9]. Hence, some Learning System offers navigational support (i.e. overviews, concept maps, link suggestions, etc.). These supports are claimed to reduce disorientation and cognitive load, but still have gaps [6], [10], [11]: (1) they are generally intended to all learners' profiles (not personalized for each individual learner) (2) make learners passive and (3) limit their free navigation. Finally (4) they draw their attention to the global structure at the expense of attention to the local structure of the text/content. That's why, before being able to remedy and support learning, it is advisable to examine a number of aspects and to bring to it an element of precision. It is well to ask: who are the learners in a situation of disorientation to be able to help and support them? What are then the consequences and the real causes of this experience disorientation?

Indeed, we consider that answers to these questions can provide very significant information on the progress of learning, even on its effectiveness. This is what we propose exploiting in this article, through an implicit modelling method, based on the collection and interpretation of interaction traces that allow enriching the learner model by labelling the kind of path. We also propose to introduce the taught domain ontology to correlate information among the various educational content viewed by the learner.

The paper is organized as follows: in section 1 we present some related work that could raise this kind of observation. Section 2 is devoted to showing some metrics to measure the disorientation. In section 3, we look at some tools of the semantic web, particularly ontology. In section 4, we develop the proposed approach to detect and evaluate the path of a learner (disoriented or not). Finally, in section 4 we present some results, which we have achieved through an implementation of our approach on a considered domain.

II. RELATED WORK

Disorientation is defined as a situation in which due to cognitive overload and to the complexity of the environment, learners do not know where they are in the network where they want to go next in the network, and how they can get there [12], [13]. In other words, disorientation means that learners are unaware their places within the whole structure and do not know how they have reached that place and how and where they will go. Likewise disorientation is also known as looseness in related literature, but the disorientation is more frequently used. The reason is that the word 'disorientation' has a more comprehensive meaning [14].

When studies conducted to decrease or overcome the problem of disorientation are examined, it's noticed that these studies generally focus on the coherence of reading in relation with the learner's level. Coherence can be defined as, ''*the degree to which the reader's navigation path (or reading order) follows a coherent line of arguments or ideas*'' [15].

Research comparing numerous studies which looked at domain knowledge and interest as variables impacting comprehension has found that learners with more subject-matter knowledge are better prepared to process or navigate digital text and therefore are apt to increase their domain knowledge, whereas those without a sufficient base of knowledge tend to struggle and thus fall increasingly behind [16]. Further, prior knowledge can help support comprehension of information accessed in digital environments by assisting readers to follow more coherent reading sequences as well as reduce feelings of disorientation [6].

Hence, in [17] the authors observed that the learners with low level of knowledge made more jumps from a subject to another one while the learners of higher level of knowledge, pursued more complete and deep explorations of the same theme. Ford and Chen [18] observed positive correlations between the previous knowledge and the more elaborate browsing behaviour (fewer reviewed pages, deeper exploration, less demand of assistance) and the speed in the realization of the learning task.

In a qualitative study led by 8 participants, Rezende, and de Souza Barros [19] have identified three profiles of navigation depending on the prior level of knowledge: "*organized navigation* " for the learners of the highest level of knowledge*,* " conceptual navigation " for the learners of an intermediate level and *"disorientated navigation* " for the learners of low level of knowledge. Other works, more qualitative, go in sense of navigation which is guided by the knowledge by showing that a lack of knowledge leads learners to pursue little developed strategies (methodical and exhaustive strategies); while a high level knowledge promotes the use of less erratic strategies [20].

In all these studied works, we could infer that it is easier to recognize a performance learner with low knowledge through the form of the path established during his learning. Separately and in combination, disorientation and low knowledge have an impact on the poorer performance in learning tasks. As matter of fact, the disorientation is considered as serious and frequent navigation problem faced learner, since most of the Learning System are delivered with web systems. Finding students' disorientation levels and preventing it, is important to improve learning. For that reason, correct disorientation measurement is obvious and inevitable.

III. DISORIENTATION MEASURE

In the literature, there were several manners which allow measuring the disorientation level. They are classified under two main topics. *The qualitative method* which is a subjective method in nature, this method consists in collects of the user opinions, through several Likert-type scale questions. Although it's simple and fast to administer, it is limited, the researcher can never know if the user is really conscious of his disorientation [21].

The second method, which is *quantitative,* bases mainly on the learner log navigation (trace). We find, for example the time spent to accomplish a task (reading page). Herder [22] found that user disorientation in hypermedia could be characterized by combining an index of reviewing behavior with median page view times. The number of cycles is also seen as a good predictor of disorientation, according to a study conducted by Juvina and Oostendorp [23]. Other works use the measure of similarity between the learner navigation trace and optimal path to detect the disorientation [8], [24]. So, for these works as much as the learner pattern was similar to the optimal pattern these indicate a coherent reading and show lesser disorientation. In contrast, indicate a weak performance and disorientation.

Indeed, our study is focused on the development and use of a new method to measure disorientation which has in fact semantic nature. According to studied work, disoriented learner made an extra effort to develop a clear idea of the relations among the pages/links and their position in the environment, have erratic strategies, made surface exploration and more jumps from a subject to another. Based on these behaviors, we have concluded that the pages which follow each other in the navigation of the disoriented learner have less relation, which means less similarity.

Our work joins on this optic, by trying to observe the form of the learner navigation path through the application of semantic similarity measures between the concepts of pages that follow each other in the navigation. That means, if the semantic distance between the concepts of pages which are following in the navigation is large, this translates an incoherence of reading and disorientation for the learner. Note that the concepts of the taught domain are well-modelled through the domain ontology of the given course, as we describe it in what follows.

IV. ONTOLOGY

The Semantic Web [25] is an understandable and navigable space by both human and software agents. It introduces an additional meaning to the navigational data of the classical web, based on a formal ontology and controlled vocabularies through semantic links. In standpoint of e-learning, it can help learners to locate, access, querying, processing and evaluating learning resources across distributed heterogeneous network, or assist teachers in creating, using, locating, or sharing and exchanging learning objects. Ontology [26] includes a set of terms, knowledge, including vocabulary, semantic relations, and a number of logic-inference rules for some particular domain. The ontology applied to Web creates thus the Semantic Web [27].
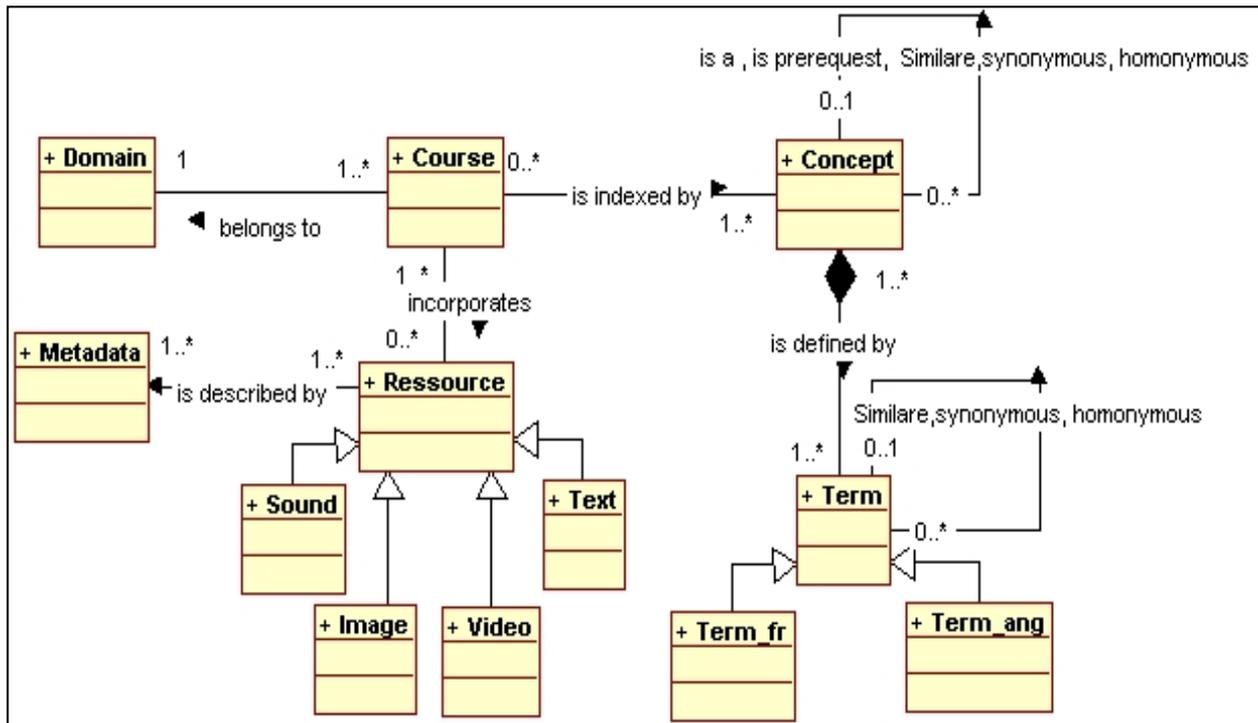
Figure 1.   The model of domain ontology [30]

Ontology [28] facilitates the sharing and reuse of knowledge, i.e. a common understanding of diverse content by persons and machines.

In our instance, the use of domain ontology resides essentially in the conceptual indexing of learning resources to facilitate the identification, research and semantic reuse of these resources.

This ontology also represents the structure of the learner's model, since it is part of the domain model, i.e. the domain ontology in our case. In our case of study, we consider that ontology is composed of a set of concepts and relations between these concepts. A unique identifier is assigned for each concept; these concepts are labelled with one or several terms. The domain ontology model that we propose is shown in Figure 1. The considered educational resources are described by a set of metadata (LOM) [29].

V.   METHODOLOGY

The approach we propose is based on the analysis of trace files of the learners in a situation of learning during their navigations on a course concerning a given topic (Figure 2). We suppose that this file contains essentially the URLS of the visited pages as well as their dates of access.

This trace file will permit us to bring back the learner navigation routes, to realize the semantic indexing of the visited pages, and finally to compute the indicator of disorientation. We detail these steps in what comes after.

A. Path of navigation

In traditional learning, interactions between the learner and the teacher are manifold, through educational materials, gestures and words. The teacher can, according to his observations, change the course of his career to suit the different profiles of its learners [31].
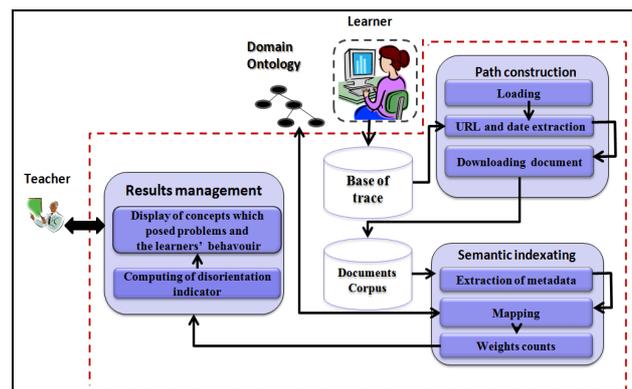

Figure 2.   Architecture of the proposed solution

In distance learning, these observations are derived from collected traces (log file). Jermann [32] defines a digital trace as a set of observations on the interaction of the learner with a system. It is defined as an observable time sequence of the browsed pages modelled by the URL and a set of actions on these pages. The collect on the client side provides data relating to the path of the learner, whether inside the learning platform or outside of it, i.e., browsing the course or the Web, communicates via forum or chat, running various applications on the local machine. We are interested in our case, particularly on the browsing of the course.

The data collected on the client side is called log files or the track files. Voluminous and very meticulous that are, it is difficult or impossible to understand them as they are by humans. It is therefore, indispensable to perform some processing on it, to make it interpretable by the cleaning process. The cleaning consists in filtering the insignificant and superfluous data from the track file. In our context, we keep only the URL of the course content and the date of the reading of this URL.

After the log file processing step, it's clear that it is possible to represent the path of the learner as a graphical record of visited pages (URL), in which the behaviour of the learners is translated by the various types of path. Indeed, we can represent a learning path for a time T by G (V, E) such as V represents the set of the visited pages and E, as a set of links activated by the learner and connecting the various pages browsed during his learning.

V = {P1, P2…. Pn}, E = {I₁,₁, I₁,₂… Iₙ, ₘ}.

Where $I_{i,j}$: indicates a link connecting page Pi to page Pj.

Note that the links that allow to pass from a page to another one, can be explicit by means of a hypertext or an hypermedia link from the page Pi, as they can be implicit if for example the learner decides to type directly the URL of page in the browser bar or access on a link if it is already marked as a favourite... etc.

The path, thus established, may contain the same pages several times (Number of cycles), since the learner can afford to revisit several times the same page.

Example1:

We suppose the following pages constitute the structure of a given course:

Course = {P1, P2, P3, P4, P5, P6}

And let us propose this path for a given learner: P1→ P2→ P4→ P2→ P3→P2→ P4→ P5

We note in this example (figure 3) that the page P2 and P4 were visited more than once by the learner and we count 2 cycles (Edges with thick dots form the first cycle, where those with the thin dots represent the second cycle)

After having the various pages browsed by the learner through the trace file, we then precede on the indexing of these browsed pages with the domain ontology.

B. Conceptual indexation

We suggest using domain ontology in order to index the browsed documents. This ontology should be defined by the course designer. The indexing process is performed in four main steps: (1) the representation of the consulted documents (2) the association of concepts to the document keywords (3) disambiguation and (4) weighting of concepts.

1) Representation of documents.

We represent every visited page by its keywords (through metadata). The keywords represented by the metadata are assigned to the documents in an equitable way, whereas their frequency of appearance, allows giving importance of a keyword compared to another. The vectorial model is the most suitable in the proposed approach for an effective representation of the documents. Each document is identified by a vector of n dimensions where each dimension corresponds to the various terms. In each term vector of a document Di, a weight is assigned:

Di= (P₁ᵢ, P₂ᵢ, …. , Pₙⱼ). Where Pⱼᵢ is the weight of the *Jth* term in the document *i*.

This weight is the function of the term frequency, collection and factor of normalization, *tf*idf* (*Term Frequency*/*Inverse Document Frequency*) [33], [34].

$$\text{TfIdf}_{i,j} = tf_{i,j} \times idf_j \quad (1)$$

$$idf_j = log\frac{N}{f_j} + 1 \quad (2)$$
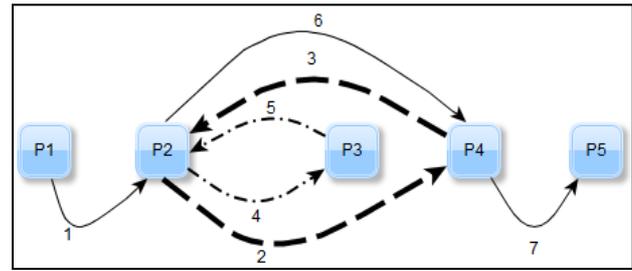

Figure 3.   Example of path navigation

Input: *document D.*

Output: *Vector of all ontology concepts belonging to keywords in document D.*

Procedure

*Let **wi**be the next word to analyze in the document d. We define the context **sent₁** which is the sentence in document D that contains the word occurrence **wi** being analyzed.*

Compute Vi = *{C1 , C2 ,..., Cn } the concepts of the ontology entries containing **wi**.*

Each *C j □ **Vi** is represented by a multiword or mono-word term.*

Rank *concepts Cj in set **Vi** in where: | C(1) | > | C(2)|> | | ... > | C(n) | // | | denotes the concept length, in terms of the number of words in the corresponding terms.*

For each *element Cj in Vi do*

Get *common words between sentᵢ and representative term of Cj , which is the intersection*

*N = ∩(**sentᵢ** ,C j)*

    *If |N | < | C j | **then** The concept-sense is not within the context.*

EndIf

*If |N | = | C j | **then** The concept-sense Cj is within the context **sentᵢ**.*

Add Cj to the set of vectors' element (index) associated to the document D.

EndIf

EndFor

Figure 4.   The algorithm of Words Mapping into Concepts [36]

Where $tf_{i,j}$ is the frequency of the term *j* in the document *i, N* is the total number of documents, f_j is the number of documents in the corpus containing the term *j*.

Once we have the keywords of each document (the vector of terms) and their weight, we proceed to their projection on the domain ontology.

2) Concept Identification.

The purpose of this step is to identify the ontology concepts that correspond to the keywords of the document. The extraction of concepts is based on the analysis of each term in the document (keyword) with all the ontology entries [35]. The concepts are referenced in the documents with simple or compound keywords. The identification algorithm is well-explained in Figure 4.

The operation of conceptualization of the terms consists of assigning to every term **t** describing the document a concept **C** of the domain ontology. However, we risk finding a term **ti** which can have **n** possible concepts in the ontology (several entrances to the ontology). As a result, the disambiguation step is then necessary.

3) Term Disambiguation.

Initially, we use the domain ontology to retrieve the various possible senses for an ambiguous term **ti** (candi-

date concepts). Then, we will examine an unambiguous term tk in the document, identified by a concept Ck in the ontology and which is linked to different candidates' concepts.

We shall then calculate a score for each candidate concepts, this score is based on the similarity measures between the various senses (candidate concepts) and the concept Ck, using measures such as measures of *Leacock and Chodorow* (or Lch) [37] *Lin* [38]and *Resnik*[39].

In our case, Lin's method is used to measure the semantic similarity. This method is one of the most theoretically studied and takes into account the information shared by the two concepts as *Resnik* but also about what distinguishes them:

$$sim(ci, ck) = \frac{2 * log\, P(mscs(ci,ck))}{log\, P(ci) + log\, P(ck)} \qquad (3)$$

Where.

- P (c) is the probability of encountering an instance of the concept c.
- mscs (Ci, Ck) is the common subsuming between the two concepts.

The candidate concept Ci with the highest score is then selected to represent the ambiguous term ti.

4) Concept Weighting.

The extracted concepts are weighted according to a method more general than *tf* *idf* named *Cfc * idf* (concept-frequency-inversed document frequency).

In this method, each extracted term necessarily represents a concept of the ontology since we used the ontology to identify them. For a concept C its frequency in document depends on the frequency of the term itself [40]. It's a ratio which is calculated as follows:

$$idf_c = log\, \frac{N}{f_c} + 1 \qquad (4)$$

$$Cf_c = \sum_{tm \in t(c)} tf_{tm} \qquad (5)$$

Where: t(c) is the set of terms corresponding to the concept C and tftm is the frequency of term t(c) in a document i. The weight of each concept in a document d is so calculated as follows:

$$Cfidf = Cfc \times idfc \qquad (6)$$

C. Computing of the disorientation indicator

After the conceptual indexation, each document (visited URL) is represented by a vector of weighted concepts. We proceed then to calculate the semantic distance between two vectors V1 and V2 corresponding to two pages that follow in the learner's navigation.

To do this, we opted for the *Jaccad index* [41]. The Jaccard similarity measure is defined by the quotient of the number of common objects by the total number of objects from which is subtracted the number of common objects:

$$(7)$$

$$sim(V1, V2) = \frac{Nc}{N1 + N2 - Nc}$$

Where N1 and ... s of vectors V1 and V2 and Nc the number of common concepts between the two vectors. The result of the similarity measure is a normalized value included in the interval [0, 1] and similarity between V1 and V2 is evoked when the score is closed 1.

The index of disorientation corresponds to the average of the distances of all the browsed pages:

$$des = \frac{\sum_{i=0}^{n-1} sim(Vi, Vi+1)}{n-1} \qquad (8)$$

Where, n represents the number of pages browsed by the learner.

The measurement result of disorientation is still a normalized value in the interval [0, 1]. Consequently, as long as the measure is big and close to 1, means they are similarity between the browsed pages that follow each other in navigation and indicates a coherent navigation path, otherwise (the measure is close to 0), interprets a dissimilarity and incoherent path, thus a learner is in trouble and needs help and monitoring. A threshold α between 0 and 1 will be fixed through experiments to effectively identify learners in a state of disorientation. So, it's good to precise that as much as similarity is large the semantic distance and the disorientation are minimal.

D. Statistical Indicators

We add to the disorientation indicator, which has a semantic character, other statistical indicators (quantitative navigation metric) that allow us to recognize, in a little way, the parts of the course which block learners and put them in disorientation. Indeed, having more than one indicator related to the learner navigation will provide us relevant information even on the causes and the consequences of the disorientation. The metrics features are:

- The frequency of page read since the beginning of the session.
- The time spent on the page (weighted average time spent per page in the session).
- Score of diversity, which is the different number of the read pages, divided by the number of the course pages.
- Score of depth, which indicates if the learner has done a deep exploration of the content of the course. It is computed by dividing the number of links activated by the learner to the total number of links existing in the course.

Example 2:

If we take the sequence of pages visited in the Example1

- The visit frequency: P1 (1 time), P2 (3 times), P3 (1 time), P4 (2 times), P5 (1 time), P6 (0 times)
- Diversity = 8/5= 1.6
- Depth=5/6=0.83. (If we suppose there are 6 links on the 6 pages, and only 5 links are activated corresponding to 5 visited pages).

VI. EXPERIMENTATION

A. Initialization of the experiment

1) Participants

To validate the effectiveness of the indicator developed in the previous section, an experiment was carried out in the undergraduate course of "PHP" with the participation of 182 students of the second year for one month. The experiment was conducted in sessions at the computer lab of the *Mouloud Mammeri University of Tizi-Ouzou* (Algeria).

2) Web-based learning system

The course is presented to the students via a standard web-browser (figure5).This system provides multiple navigation tools, which include the main menu, keyword search, hierarchical map and alphabetical index. The keyword search is to allow learners to locate information for particular concepts with a search box. These navigation tools serve different purposes, so learners are allowed to have a freedom to develop their own navigation strategies.

The web-based learning material was introduced to the participants in the first week and was available for the students from then on, for an overall four week time period. A reference manual is available in a form of video and PDF document.

When students logged into the system, they were provided the tasks of the week and were told to complete them within one course-session. Then, they met the task and worked on the contents and examples in any sequence on the web-based learning material, which contains 81 pages. Students' interaction logs were stored in the database to compute disorientation. At the end of the experiment, an exam of 20 questions was made using a *multiple choice* format, for all the learners.

Hence, the PHP ontology is mainly composed of 8 top concepts and 49 sub-concepts, the main concepts are: (C1: Variable and Constant, C2: Operator, C3: Control Structure, C4: Function, C5: Management Form, C6: MySQL, C7: Treatment on files, C8: Session and Cookie).

B. Data Analysis

The aim of this study is to detect the learners in disorientation behaviour and investigate consequences and causes. In order to reach our aim, learners score obtained from the test are worked as attributes to create clusters.

Different clustering algorithms have been used to group students, such as: hierarchical agglomerative clustering, a clustering algorithm based on large generalized sequences to find groups of students with similar learning characteristics based on their traversal path patterns and the content of each page they visited[42]; an improvement in the matrix-based clustering method for grouping learners by characteristics in e-learning [43]; a K-means clustering algorithm to effectively group students who demonstrate similar learning portfolios (students' assignment scores, exam scores and online learning records) [44].

Among a variety of clustering techniques, K-means algorithm is widely used to partition the data into several clusters according to their similarities [45]. The suitable number of clusters is determined based on, not only the smallest distance between the features in a same cluster but also the largest distance between the features in different clusters. After doing so, we found that the K-means algorithm produces more efficient outcomes for three clusters.

C. Evaluation of results

The result of the observations consists of 182 trace files corresponding to each learner's interaction (182 learners). The analysis and treatment of trace files allow us to obtain, an additional corpus which merging all documents visited by all learners, in total we will have 183 corpus of documents. The analysis of these corpuses allowed us to calculate 182 disorientation values. Same is the number of the score test (questionnaire).

Figure 5. The Web-based learning system

So, the three clusters were created with a K-means algorithm. As shown in Table 1, the percentage of learners within each cluster is satisfied because the number of members in each cluster is reasonably balanced. The Mean and standard deviation (SD) of the score test for each cluster are shown in Table 1. The tendency of each cluster is shown below:

TABLE I.
SCORE OF TEST FOR EACH CLUSTER

Cluster	Mean	SD	Instances
C1	13.66	6.77	23%(42)
C2	10.22	4.64	45%(82)
C3	7.54	3.25	32%(58)

- Cluster 1(C1, N=42): students in this cluster have responded well to the test and have generally a good performance; their results vary from 12/20 to 17/20.
- Cluster 2(C2, N=82): the students of this cluster are average, as indicated by their result in a test, which are in the interval of 9/20 to 11/20.
- Cluster 3(C3, N=58): group students who had a bad performance, which diverge between 5/20 and 8/20.

After getting the clusters according to the evaluation score, we calculate in the following the Mean and SD of the disorientation score for each cluster:

TABLE II.
SCORE OF DISORIENTATION FOR EACH CLUSTER

Cluster	Mean	SD	Max	Min
C1	0.81	0.42	0.85	0.72
C2	0.56	0.23	0.68	0.45
C3	0.33	0.09	0.39	0.24

- Cluster 1(C1, N=42): we observe that the disorientation score of students in this cluster was the highest; this means a good similarity between the viewed page of their path. The disorientation score of the student who has the best performance (17/20) is 0.76.
- Cluster 2(C2, N=82): the students of this cluster have average similarities between the page which follow each other in the navigation, as indicate by their disorientation score which is also average (0.56).
- Cluster 3(C3, N=58): this cluster is the one of the three other clusters which has the least value of the disorientation indicator (0.33), this explains as well the failure of the learner in the test with the lowest test score (7.54). The week student (5/20) of this cluster has 0.29 as disorientation score.

Thus, it is suitable to fix the disorientation threshold α as the mean of the Minimum value of the cluster 2 (0.45) and the Maximum value of the cluster 3 (0.39). As the result, the thresholds α is fixed at 0.43.

At the present, it remains to shed light on the consequences and causes of disorientation. To do this, we propose to analyze the mean and the SD of the frequency browsed page, the diversity and the time on the reading a page for each cluster. These are well detailed in the following tables:

TABLE III.
SCORE OF FREQUENCY PAGES READ, DIVERSITY, DEPTH AND READ TIME FOR EACH CLUSTER

Cluster	Frequency of pages read		Diversity		Depth		Reading time	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
C1	1.21	0.42	1.33	0.31	0.95	0.12	5	2.7
C2	1.90	0.73	1.78	0.64	0.79	0.09	6	2.6
C3	2.82	1.21	3.15	1.81	0.52	0.05	2	1.1

We detect that students of a cluster (C3) who have disorientation (des=0.33) behaviour, have a superficial exploration of the course content (depth= 0.52) and an attitude to revisit some pages of the course more than once (2.8), which means a large consultation of the concepts contained in these pages. As a result, the diversity is large (3.1), since it is also influenced by the number of read pages. Whereas for the mean of the reading time, in the cluster C3, learners spend the least time (2 minutes) than the two other clusters (6 minutes for C1 and 5 minutes for C2).

Therefore, we deduce that among the consequence of disorientation which we have raised, we found at first, the weak learner performance, secondly, we observed that the learner tends to make a cyclical navigation path which is inferred by some pages revisited more than once, finally the weak effort of learning is interpreted by the short time that they take to read their pages course and their superficial exploring of content.

To recognize the main causes of the learners' behaviour disorientation, in our case of study, we have based solely on the analysis of the students results of the cluster (C3), which have the number of 58 students. However, we have listed the pages whose visits' frequencies are greater than one consultation, which are presented in the following graph, and in addition compared with those of the other clusters.

The pages that have been presented in the graph above were focused especially on the concepts and sub-concepts of *Management of Form* (page 39, 40 and 41), *Mysql* (55, 56) and *session* (68, 69, 70 and 73). As we observe, the pages that correspond to the concept of forms were revisited more than once likewise the two other clusters, this is argued by the complexity of this concept at first and on the height relationship to the *session* and *MySql* concepts at second.

The graph in next (figure 7), summarizes the mean score obtained by learners for the three concepts and their sub-concepts from the different clusters.

We remark that the score obtained by the students of the cluster 3 for the three preceding concepts are the lowest (Score <0.4), this justifies clearly the high frequency of pages read that articulate about these concepts (Figure 7).
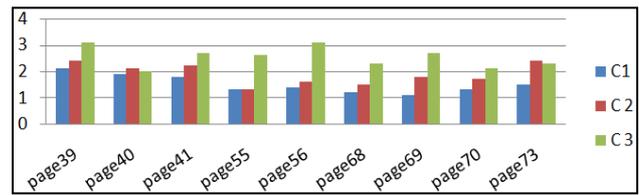

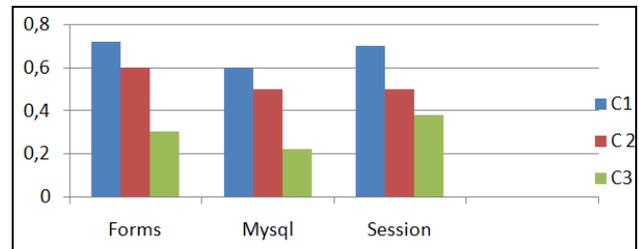Figure 6.   The frequency of the read pages

Figure 7.   The test score of the three most-consulted concepts

In summary, learners in Cluster 3 show worse learning performance than those in the other two clusters. As mentioned in the above section, learners in Cluster 3 also have disorientation problems than those in the other two clusters. It implies that the learners with more disorientation problems do not perform as well as those without disorientation problems. In general, the problem of understanding is often begun from the concepts contained in the pages revisited than twice and has a low semantic similarity with those that follow, which interpret that there are a bad definition and explanation of concepts and/or the overload of information on these pages.

D. Discussion

This experiment allowed us at the first to set some parameters such as disorientation threshold fixed at 0.43, the average time of reading page at 5 minutes, and the average diversity at 0.7.

It has also enabled us, to check the hypothesis, which we have highlighted about the fact that the disorientation is generally started from the pages whose content is poorly assimilated by learners. These were detected through observation of learners in cases of disorientation and who have a tendency to revisit the same pages repeatedly, so they establish a circular or ring navigation route. The learning difficulties encountered by these learners are often related to the concepts enclosed in these revisited pages.

We have also shown the apparent consequence of disorientation on the learning level, the correlation is so positive (r=0.281; p<0.05), thing that was always checked in all research work. Indeed, disorientation is one of the major keys of failure in learning. Another result, that we have raised, is related to the learning effort, which we have translated in our case first, by the modest time undertaken by the learner to attend his course, the correlation is also positive (r=0.465; p<0.05). Secondly, by the superficial exploring of the content with positive correlation (r=0.337; p<0.05).

As for the disorientation causes, always in our case of study, we have underlined some concepts that were problematic for learners, and which are present in some pages that have been consulted several times. In fact, the relationship between disorientation and a page reading frequency is negative (r=-0.372; p<0.05). This is often due to

the poor of structuring and designing of the course, as well as overload of the learner cognitive and the information contents in the pages of the course.

To this end, there is a need to remove the causes of disorientation problems, so that the consequence can also be avoided. Hence, the proposed approach facilitates the task either for the tutor to monitor the learner in order to identify disoriented learners in a timely manner and for the author of the course, to review the course on pages (concepts) that are problematic and restructured them in a way to respond to the learners' needs. These Two procedures, whether from the tutor or the course author, are processes to promote the training and help learner by the adaptation of the contents as well as ensuring a good follow-up.

VII. CONCLUSION

Through our study, we identified the need to observe the navigation behaviour of a learner in order to detect the type of path that he made during his apprenticeship. This path is different from a learner to another according to the need and the difficulties that he met during its learning. Detecting the type of behaviour may well help us to recognize learners in difficulties because an unstructured and inconsistent path requires more intent from the tutor to assess the learner to provide him assistance and by the course designer for revision and restructuring of the course.

For this, we mainly interested in this paper to the disorientation indicator. This is based on the semantic distance between each pair of pages that follow in the learning navigation. The computing method of the disorientation indicator has semantic nature as it was detailed in this paper, and reliability of this indicator has been well-validated through an experimentation which we conducted for 182 second year students. The experiment has allowed us to fix certain thresholds, such as disorientation, that let us recognize one of the causes of disorientation, which is the cognitive overload, interpreted in some pages of course which are detected through the high frequency of reading page. Our perspective is based on the integration of several other indicators to better assess and understand his condition and behaviour.

REFERENCES

[1] J. Waniek, and T. Schafer, "The role of domain and system knowledge on text comprehension an information search in hypermedia," *Journal of Educational Multimedia and Hypermedia* vol. 18, no 2, pp. 221-240, 2009.

[2] A. Calcaterra, A. Antonietti, and J. Underwood, "Cognitive style, hypermedia navigation and learning," *Computers and Education*, vol. 44, no 4, pp. 441-457, 2005. http://dx.doi.org/10.1016/j.compedu.2004.04.007

[3] F. Amadieu, A. Tricot, and C. Mariné, "Individual differences in learning from hypermedia: Learners' characteristics to consider to design effective hypermedia," *6th International conference on human system learning* – ICHSL, 14–16 May. Toulouse, 2008, pp. 1-9. Toulouse: IEEE France Section.

[4] P. Brusilovsky, and N. Henze, "Open corpus adaptive educational hypermedia. In The adaptive web: Methods and strategies of web personalization," edited by P. Brusilovsky, A. Kobsa and W. Neidl, *Lecture Notes in Computer Science*, 2007. Vol. 4321, Berlin Heidelberg New York: Springer-Verlag, pp. 671-696.

[5] S.A. Sahraoui, "Problèmes d'intégration des Hypertextes/ Hypermédias dans l'apprentissage des langues. Synergies," Algérie. 2007. pp. 251-257.

[6] F. Amadieu, A. Tricot, and C. Mariné, "Interaction between prior knowledge and concept-map structure on hypertext comprehen-

sion, coherence of reading orders and disorientation," *Interacting with computers*, vol. 22, no 2, pp. 88-97, 2010. http://dx.doi.org/10.1016/j.intcom.2009.07.001

[7] T. J. F. Mitchell, S. Y. Chen, and R. D. Macredie, "Cognitive Styles and Adaptive Web-based Learning," *Psychology of Education Review*.vol. 29, no 1, pp. 34-42, 2005.

[8] T. Güyer, B. Atasoy, and S. Somyürek, "Measuring Disorientation Based on the Needleman-Wunsch Algorithm," *The International Review of Research in Open and Distributed Learning*, vol. 16, no 2, April – 2015.

[9] J.F. Rouet, "Interactivité et compatibilité cognitive dans les systèmes hypermédias," *Revue des Sciences de l'Éducation*, 1999. http://dx.doi.org/10.7202/031994ar

[10] R. D. Ruttun, and D. Robert, "Macredie the effects of individual differences and visual instructional aids on disorientation, learning performance and attitudes in a Hypermedia Learning System," *Computers in Human Behavior, vol.* 28, pp. 2182–2198, 2012. http://dx.doi.org/10.1016/j.chb.2012.06.026

[11] S. A. Sullivan, S. Puntambekar, "Learning with digital texts: Exploring the impact of prior domain knowledge and reading comprehension ability on navigation and learning outcomes," *Computers in Human Behavior,* vol. 50, pp. 299–313, 2015. http://dx.doi.org/10.1016/j.chb.2015.04.016

[12] J. Conklin, "Hypertext: An Introduction and Survey," *IEEE Computer*, vol.20, no 9, 1987, pp. 17-41, 1987.

[13] Y.L. Theng, C. Rigny,H. Thimbleby, and M. Jones, "Cognitive task graphs and executable user models for better hypertext," *APCHI'96*, pp. 421 - 433, Singapore.

[14] M. Firat, and A. Kuzu, "Semantic web for e-learning bottlenecks: disorientation and cognitive overload," *International Journal of Web & Semantic Technology* (IJWesT) vol.2, no.4, October 2011.

[15] R. I. Madrid, H. Van Oostendorp, and M. C. P. Melguizo, "The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order," *Computers in Human Behavior*, vol.25, no 1, 66–75, 2008. http://dx.doi.org/10.1016/j.chb.2008.06.005

[16] P. A. Alexander, J. Kulikowich, and T. Jetton, "The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts," *Review of Educational Research*, vol.64, pp. 201-252, 1994.

[17] E. Carmel, S. Crawford, and H. Chen, "Browsing in hypertext: a cognitive study," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no 5, 865-884, 1992. http://dx.doi.org/10.1109/21.179829

[18] N. Ford, and S. Y. Chen, "Individual differences, hypermedia navigation, and learning: an empirical study," *Journal of educational multimedia and hypermedia*", vol. 9, no4, pp. 281-311, 2000.

[19] F. Rezende, and S. de Souza Barros, "Students' navigation patterns in the interaction with a mechanics hypermedia program," *Computers & Education*, vol.50, no 4, pp. 1370-1382, 2008. http://dx.doi.org/10.1016/j.compedu.2006.12.011

[20] P. Mishra, and A. Yadav, "Using hypermedia for learning complex concepts in chemistry: A qualitative study on the relationship between prior knowledge, beliefs, and motivation," *Education and Information Technologies*, vol. 11, no1, pp. 33-69, 2006. http://dx.doi.org/10.1007/s10639-005-5712-6

[21] J. S. Ahuja, and J.Webster, "Perceived disorientation: An examination of a new measure to assess web design effectiveness," *Journal of Interacting with Computers*, vol. 14, no 1, pp. 15-29, 2001. http://dx.doi.org/10.1016/S0953-5438(01)00048-0

[22] E. Herder, "Revisitation patterns and disorientation," In *German Workshop on Adaptivity and User Modeling in Interactive Systems* ABIS03, 2003, 291-294. Karlsruhe. Germany: University of Karlsruhe.

[23] I. Juvina, and H. V. Oostendorp, "Individual differences and behavioral metrics involved in modeling web navigation," *Universal Access in Information Society,* vol. 4, no 3, pp. 258–269, 2006. http://dx.doi.org/10.1007/s10209-005-0007-7

[24] P. Dias, and A. P. Sousa, "Understanding navigation and disorientation in hypermedia learning environments," *Journal of Educational Multimedia and Hypermedia*.

[25] D. Dicheva, "Ontologies and Semantic Web for E-Learning. Handbook on Information Technologies for Education and Training," *Springer Berlin Heidelberg*, 2008, pp. 47-65. http://dx.doi.org/10.1007/978-3-540-74155-8_3

[26] U. Siti, A. Rohiza, M. Shakirah, "Ontology of Programming Resources for Semantic Searching of Programming Related Materials on the Web," *Information Technology* (ITSim), *International Symposium, IEEE*, 2010, pp. 698-703

[27] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific American*, 2001, pp. 29-37. http://dx.doi.org/10.1038/scientificamerican0501-34

[28] T. Zschocke, and J. deLeon, "Towardsan Ontology for the Description of Learning Resources on Disaster Risk Reduction," CCIS 111, *Springer-Verlag* Berlin. *WSKS,2010*, pp. 60–74

[29] IMS Global Learning Consortium. "IMS Meta-data Best Practice Guide for IEEE," 1484.12.1- 2002 *Standard for Learning Object Metadata*, 2006.

[30] S. Ait Adda, and A. Balla, "The Use of Ontology in Semantic Analysis of the Learner's Queries on the Web through Search Engines," IJWA vol.6, no 3, pp. 121-130, 2014.

[31] M. Ben Sassi, and M. Laroussi, "Towards learners' tracks standardisation" *frantice.net*, numéro 5, septembre 2012. www.frantice.net.

[32] P. Jermann, A. Soller, and M. Muehlenbrock, "From mirroring to guiding: A review of state of the art technology for supporting collaborative learning," *In Proceedings of the First European Conference on Computer-Supported Collaborative Learning*, 2001, pp. 324-331.

[33] S. Robertson, and K. Sparck Jones, "Relevance weighting for search terms," *Journal of The American Society for Information Science*, vol. 27, no 3, pp.129-146, 1976. http://dx.doi.org/10.1002/asi.4630270302

[34] K. Sparck Jones, "Experiments in relevant weighting of search terms," *Information Processing & Management*, vol. 15, no 3, pp. 133-144, 1979. http://dx.doi.org/10.1016/0306-4573(79)90060-8

[35] M. Baziz, M. Boughanem, G.Pasi, and H. Prade, "An Information Retrieval Driven by Ontology from Query to Document Expansion," *Proceedings of the 8th Conference on Large-Scale Semantic Access to Content* (Text, Image, Video and Sound), RIAO, 2007, pp. 301-313.

[36] S. Ait Adda, and A. Balla, "The use of ontology in semantic analysis of the published learner's messages for adaptability," *Proceedings of the 4th International Conference of Swarm Intelligence Based Optimization -ICSIBO 2014*, LNCS, 2014, pp. 106-114.

[37] C. G. Leacock, A. Miller, and, M. Chodorow, "Using corpusstatistics and WordNet relations for sense identification," *Comput. Linguist.* 24, 1, Mar.1998, pp. 147-165.

[38] D. Lin, "An information-theoretic definition of similarity," *Proceedings of 15th International Conference on Machine Learning*, 1998.

[39] P. Resnik, "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language," *Journal of Artificial Intelligence Research* (JAIR), vol. 11, pp. 95-130. 1999.

[40] M. Dragoni, C. Pereira, and A. Tettamanzi, "An Ontological Representation of Documents and Queries for Information Retrieval Systems," *IEA/AIE 2010*, Part II, LNAI 6097, *Springer-Verlag Berlin Heidelberg*, 2010, pp. 555–564.

[41] N. Ljubešić, D. Boras, N. Bakarić, and J. Njavro, "Comparing Measures of Semantic, Similarity," *Proceedings of the ITI 2008, 30th International Conference on Information Technology Interfaces,* IEEE, 2008, pp. 675 – 682.

[42] T. Y. Tang, and G. Mccalla, "Student modeling for a web-based learning environment: a data mining approach," In *Conference on Artificial intelligence*, Edmonton, Canada, 2002, pp. 967-968.

[43] K. Zhang, L. Cui, H. Wang, and Q. Sui, "An improvement of matrix-based clustering method for grouping learners in e-learning," *In International Conference on Computer Supported Cooperative Work Design*, Melbourne, Australia, 2007, pp. 1010-1015. http://dx.doi.org/10.1109/cscwd.2007.4281577

[44] C. Chen, M. Chen, and Y. Li, "Mining key formative assessment rules based on learner profiles for web-based learning systems," *IEEE International Conference on Advanced Learning Technologies*, 2007, Japan, 1-5. http://dx.doi.org/10.1109/icalt.2007.189

[45] J. Han, and M. Kamber, "Data mining, concepts and techniques: Cluster analysis," San Francisco: Morgan Kaufmann, 2001.

AUTHORS

S. AitAdda is a PhD Student at National High School of Computer Science, Oued Smar, Algiers, Algeria , ESI (e-mail: s_ait_adda@ esi.dz).

N. Bousbia is a Doctor at National High School of Computer Science, Oued Smar, Algiers, Algeria, ESI (e-mail: n_bousbia@esi.dz).

A. Balla is a Professor at National High School of Computer Science, Oued Smar, Algiers, Algeria, ESI (e-mail: a_balla@esi.dz).