Indicators for Deducting the Learners’ Learning Styles: Case of the Navigation Typology Indicator

Nabila Bousbia¹,², Jean-Marc Labat², Issam Rebai³, Amar Balla¹
¹ LMCS-ESI, Ecole nationale Supérieure d’Informatique, BP 68M, 16270, Oued-Smar, Algiers, Algeria
{n_bousbia, a_balla}@ini.dz
² LIP6, Laboratoire de Paris 6, 104, Avenue du Président Kennedy 75016, Paris, France
{nabila.bousbia, jean-marc.labat}@lip6.fr
³ LIMSI-CNRS, Université de Paris-Sud, 91403, Orsay, France
issam.rebai@math-info.univ-paris5.fr

Abstract
Research in individual differences and in particular, learning and cognitive style, has become a basis to consider learner preferences in a web-based educational context. How learner’s learning style influences his/her navigation behavior has been investigated by several studies, which indicate that we can deduce the learning style from the navigation behavior. In this paper, we propose an indicator of “navigation typology”. We detail the way in which this indicator is calculated, based on tracks analysis, which are aggregated into low and intermediate level indicators to determine the value of the navigation typology.

1. Introduction
Over the past decade, great research efforts have been made towards the analysis of learners’ activities and interactions in a web-based educational context. They focus on the interpretation of the raw data collected during the learning session called tracks. These tracks, defined as a temporal sequence of observed actions, provide knowledge about the activity, synthesized using variables that we call “learning indicators”. This transformation process requires to know, on the one hand, the relation between the indicators and the tracks; and on the other hand, the dependences between the indicators according to their interpretation levels (low level, having no meaning alone, intermediate, and high level, interpretable, deduced from other indicators) [3].

Thus, the problem is to infer high-level information about the learner from low level ones: the navigation tracks (visited URLs, clicks, etc.).

In this research, we are particularly interested in automatically detecting the learner’s learning style (LS) [4] related to how individuals prefer to organize and represent information [13].

Earlier research [8][7][11] revealed that learning patterns, such as non-linear learning, have significant effect on students’ LS in a hypermedia learning system. Thus, we found relevant to consider the “navigation typology” indicator, describing how students navigate in educational resources.

In this paper, we describe the calculation process of the “navigation typology” indicator, through the tracks interpretation on the learner’s interactions with hypermedia courses. The courses are structured in activities and available on e-learning platforms (SCORM¹ resources).

2. Learning styles and indicators
Learning indicators are variables that indicate ‘something’ related to the mode or the ‘quality’ of individual activity, the mode or the quality of the collaboration, the process or the quality of the collaborative product [11]. They result from the

¹ Sharable Content Object Reference Model
http://www.adlnet.org
processing of one or more tracks, to give sense to the collected raw data.

To determine these indicators, we relied on the one hand, on the state of the art studies on digital tracks resulting from learning situations [10]; on the other hand, on the analysis of feedback that teachers want to have about their students. This was done through an investigation, carried out with teachers [3]. Thus, we detected a set of indicators that we classify, according to their interaction level (page, course, platform and session), and the level of interpretation they provide [6]: low level (having no meaning alone), intermediate, and high level (interpretable, inferred from other indicators) [3].

The selection of indicators, on which automatic detection of learning style is based, is a significant factor influencing the effectiveness of our approach. A critical issue is to identify: (i) which learners’ LS can be detected [4], and (ii) which actions are indicators of this style. For example, in case of the Field-Dependent/Field-Independent categorization [15], the way learners navigate is more appropriate as an indicator of their style than the specific type of material they select [8]. The latter may be valuable information about other categorizations such as Verbalisers/Imagers.

This approach is a new research direction, relatively little work exists in the literature [12]. According to [11], indicators that have been investigated for several learning style categorizations are: (i) navigational indicators (number of hits on educational resources, preferable format of presentation, navigation pattern); (ii) temporal indicators (time spent in different types of educational resources proposed); (iii) performance indicators (total learner attempts on exercises, assessment tests). In this paper, we are interested in the first type of indicators, namely, the navigation typology indicator.

3. Navigation typology indicator

Among the high level indicators we identified, we define the indicator of navigation typology, from the user modeling oriented works interested in interpreting the navigation motives of the Web users. In these studies, it is often referred to the Canter’s five classes [2][5]: scanning, browsing, searching, exploring, and wandering. From this taxonomy and tests we have done in our educational context, we defined 4 values for our indicator:

- **Overviewing**, this value is close to the Canter’s “scanning” value. It implies that the learner is covering a large proportion of pages constituting the course. Through this fast-reading, the user seeks to acquire a global “panoramic” view of the course.

- **Flitting**, close to “wandering”. It is a journey without a strategy or a particular goal. The main difference with the overviewing typology is the lack of focusing on the course.

- **Studying**, corresponds to a partial or complete reading of the course pages, with a span of time on each.

- **Deepening**, is rather close to the preceding value. It describes a learner who remains a relatively long time on a course, careful with details, and possibly seeking for Web documents related to the course topics.

Getting such information is not easy because the user’s navigation is versatile. For example, a learner, browsing a course, may suddenly decide to look for precise information.

Therefore, from the proposed definitions for the four values of the navigation typology, we have found that it is necessary to identify: (i) the consultation duration of the course pages to distinguish a rapid reading of pages of the course from a complete one, (ii) the importance of the proportion of visited pages to identify the level of details achieved, (iii) the course browsing pattern to distinguish an overviewing or flitting from a searching activity, and (iv) the semantic link between the visited pages to know whether the learner focuses on the same topic or not.

Thus, we have selected 5 intermediate indicators to find these values. Their calculating methods are presented hereafter. However, it should be noted that we consider a context of a course without an imposed scenario and which has a tree structure. It includes several activities that can include sub activities. Leaf activities point to the resources (hypermedia content, in particular HTML pages).

3.1. Intermediate indicators calculation

3.1.1. Intermediate duration of a page consultation.

It is the total duration (\( D_{pdc} \)) divided by the number of different visited pages (\( N_{pdc} \)):

\[
D_{mc} = \frac{D_{pdc}}{N_{pdc}} \tag{1}
\]

To compute the total duration of the course consultation, we calculate the sum of the consultation durations of the children activities of the course node, according to the formula (2):

\[
D_{cour} = \sum_{i=1}^{N} D_{act} \tag{2}
\]

\[
D_{act} = \sum_{i=1}^{K} D_{node} \tag{3}
\]
\[ D_{\text{cours}}: \text{Consultation duration of a course}; \]
\[ D_{\text{Acti}}: \text{Consultation duration of an activity } Acti; \]
\[ N: \text{The number of children activities of the course node}; \]
\[ K: \text{The number of the visited nodes of the activity}. \]

It should be noted that we have two kinds of activity: the leaf activity, their children nodes are the visited pages, and the internal activity, their nodes can be other activities. It is also necessary to note that the consultation duration of a page is the effective duration, in other words, during calculation, we sum the durations between two actions with the subtraction of the non activity duration\(^2\).

### 3.1.2. Consultation type

This indicator can have three values: *surface*, *average* or *systemic*. It determines whether the learner deepens his/her consultation of the course or not [1]. However, it is necessary to take into account the course structure. In our case, we consider a tree structure; the ideal is to manage to consult the maximum of sheets. To calculate this indicator, we calculate initially the consultation type on the level of the leaf activities, and then at the level of the parents activities to find its value on the course level.

#### 3.1.2.1. Consultation type on the level of a leaf activity

It is the ratio of the number of consulted pages of the activity to the total number of pages associated to this activity:

\[
Tc_a = \frac{N_{pca}}{N_p}
\]  

\( Tc_a \) is the consultation type indicator on the leaf activity level; \( N_{pca} \) is the number of consulted pages of the activity; \( N_p \) is the number of pages constituting the activity.

The value of this indicator is compared with the thresholds \( 1/3 \) and \( 2/3 \). These values were selected after tests. Consequently:
- If \( Tc_a > 2/3 \), we judge that the consultation of the activity is *systemic*, \( Tc_a \) will be around to 1;
- If \( 1/3 \leq Tc_a \leq 2/3 \) we judge that the consultation is *average*, \( Tca \) will have the value \( 1/2 \);
- Otherwise (\( Tca < 1/3 \)), the consultation is in *surface*, \( Tc_a \) will be around to 0.

Consequently, for each leaf activity, we will have one value for \( Tc_a \): 1, if the consultation is *systemic*, 0,5, if it is *average*, 0 if not.

#### 3.1.2.2. Consultation type on the level of internal activity

Each child node of the activity will have a consultation type value, “1”, “0” or “0.5”. To find the consultation type on the parent node, we calculate the average of the values of its child nodes and we apply the same classification according to the three cases presented above.

### 3.1.3. Browsing pattern

In order to detect the navigation typology, then the learning style, we thought it relevant to consider the sequential dimension of the learner’s browsing path. Among the indicators studied on the filed of users’ Web browsing behavior semantics, most research uses the browsing pattern indicator proposed by Canter and al. [5]. Formalizing the browsing sessions as a graph, they distinguish 4 patterns: 1) *Pathiness*: a path that does not visit the same node twice; 2) *Ringiness*: a ring is a route that returns to the starting node; 3) *Loopiness*: a loop is a ring that does not contain other rings; 4) *Spikiness*: a spike is a route that retraces the original path on the return journey.

The recognition of these four forms is far from trivial, especially the extraction of loops and rings whose nesting can be complex. In addition, tests developed with several students have made us aware that we must take into account the nature of the visited pages, the course structure and the point of departure and return for each form. That is why we have defined three values to our “browsing pattern indicator” related to our learning context:

- *Scholar*: This value is retained when the dominant Canter’s pattern is a path with a few returns back (a few loops), or when there is a spike;
- *Star*: This form occurs when the learner tends to often return to the same node or centric nodes (many loops to one or some nodes);
- *Dispersion*: it is a ring or a mixture of forms; in other words, learners tend to move in all directions.

Furthermore, to build this indicator, it is necessary to filter the graph depending on the interaction level considered to determine the indicators on which it depends. In general, it mainly depends on the intermediate indicator \( PL \) or “Path Linearity”. To calculate this indicator, we make the ratio of the number of different consulted pages (nodes, \( N_{pdc} \)) to the number of consulted pages (number of steps, \( N_{pc} \)):

\[
PL = \frac{N_{pdc}}{N_{pc}}
\]  

---

\(^2\) **Additional data that indicates from how long without any action we consider the learner as inactive.**
The number of different consulted pages is higher than or equal to the number of visited pages. So, the value of PL is in the interval [0,1]. If the value is close to 1 (up to 0.8), the browsing path is linear. The browsing pattern indicator will take the value “scholar”. Otherwise, a recursive algorithm is applied to the browsing graph according to the interaction level considered to find the existing forms, the number of nodes and the duration of each sub graph to detect the dominant pattern.

3.1.4. Semantic proximity. The pages differ in terms of size, function, types of contents, etc. Starting from this point of view, the visit of the page is meaningful only when we can describe its thematic and / or functional content. For this and to understand what the learner is doing, it is useful to detect whether the viewed pages outside the course are related to its subject or not.

To get this information, we propose the Prox or “proximity” indicator, which allows, for a given course, to estimate its degree of proximity or similarity with the Web pages consulted in the same time interval (i.e.: consulted at the same time as the course). The calculation of this indicator is carried out with a dedicated sub-system [9]. The indicator is considered by the average of the values of similarity between the vectors of concepts (concept, weight) describing the course and each Web page:

\[
\text{PROX (Course) } = \frac{\sum D_i \times \text{Sim (Course, Pi)}}{\sum D_i}
\]  

(Course) is the studied course; Pi is a Web page consulted at the same time as the course; Di: the consultation duration of the page Pi; Sim (Course, pi): the semantic similarity between the course and the page Pi. This similarity is calculated by the following formula:

\[
\text{Sim (q, d) } = \frac{\sum q_i \times d_j \times \text{Sim(i, j)}}{\sum q_i \times d_j}
\]

i represents the concepts of the document q (the course Course); j is the concept of the document d (the page P) having the max similarity with i; q_i is the weight of the concept i in the document q; d_j is the weight of the term j in the document d; Sim(i,j) is the semantic similarity between the two concepts i and j, calculated using the WordNet\(^3\) ontology. We chose Lin’s measure [10] considering the good performances that it provide. We propose this last formula (7) [9] following the modification of the formula suggested in [14] that calculates the semantic similarity between a document and a request. Instead of summoning all the similarity values two by two, we take only the maximum similarities.

As similarity measures between concepts/documents are always included in the interval [0, 1], the proximity indicator takes its values in this interval. It reaches the maximum value 1 when the concepts visited by the learner on the Web are identical to those of the course consulted, and 0 in the contrary case. Considering the uncertainty on the reality of the semantic proximity, we selected only two values: the proximity is strong when the Prox value is greater than 0.5, low otherwise. Note also that associate navigation duration to the pages attributes them a weight. A page that has been visited for a short time will have less influence than another which has been for a while longer. This is to reflect, for example, pages that are open as indicated by the search engine but which are irrelevant. Finally, as mentioned previously, the durations are the effective consultation durations.

3.1.5. Course duration rate Dc/Ds. Following the tests, we noted that it is useful to consider the percentage of the course consultation duration Dc compared to the session total duration Ds. This indicator allows contextualising the values of other indicators. For example, we can find a low semantic proximity, whereas the learner visits pages in connection with the course. That is due to the fact that the duration spent in the course is much greater than that spent on these pages. Consequently, we consider that the course consultation is strong if the rate Dc/Ds is higher than 50%, low otherwise.

3.2. Calculation of the navigation typology indicator

For each intermediate indicator, we determined 2 or 3 values. To find the 4 values of our indicator of navigation typology, we based them on tests and statistical studies. The thresholds are additional data provided by the user (the teacher or the analyst). However, default values are envisaged. Table 1 summarizes the deductions obtained during the tests and the values associated.

The results were calculated on a limited number of learners. In fact, we are currently preparing an experimentation which involves a greater number of students. However, these tests enabled us to notice that

\(^3\) http://wordnet.princeton.edu/
the formulas of some indicators can lead to skews. It is for example the case of the proximity indicator. The first tests have given us values going up to 0.8, but most of the values did not exceed 0.45. This is due on the one hand to the fact that the consultation duration influences the results, and on the other hand, to the fact that we use a similarity measure only based on the relation “is-a” in WordNet, a general ontology which covers several fields, but which does not contain all the terms. So, we are well aware that such information is heuristic and that the role is really to give an indication, not to provide a definitive result.

<table>
<thead>
<tr>
<th>Indicator Type</th>
<th>Dme</th>
<th>Te</th>
<th>Fp</th>
<th>Pros</th>
<th>Dc/Ds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flitting</td>
<td>Low</td>
<td>Surface 0</td>
<td>Dispersion</td>
<td>Low [0, 0.4]</td>
<td>Strong</td>
</tr>
<tr>
<td>Overviewing</td>
<td>Low</td>
<td>Average</td>
<td>Low [0, 0.4]</td>
<td>Strong &gt;50%</td>
<td></td>
</tr>
<tr>
<td>Studying</td>
<td>Strong</td>
<td>Surface</td>
<td>Low [0, 0.4]</td>
<td>Strong</td>
<td></td>
</tr>
<tr>
<td>Deepening</td>
<td>Strong</td>
<td>-</td>
<td>Strong Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table1. Deduction of the navigation typology from the intermediate indicators

4. Conclusion

In this article, we presented our approach to indicator calculation through the example of calculation of the high level indicator of “navigation typology”. This indicator intervenes in the automatic detection of the learning styles by the observation of the behavior in a web-based learning. The calculation of this indicator is not simple. It depends on several intermediate indicators, some of which are also complex to calculate. We will continue the development of our prototype in order to carry out more complete experiments. We shall report on our progress in forthcoming papers.

5. References


