

Evaluation de la désorientation de l'apprenant dans un système d'apprentissage

Samia Ait-Adda^{1,2}, Nabila Bousbia²

¹Université Mouloud Mammeri, Tizi-Ouzou, Algérie

²LMCS-ESI, Ecole nationale Supérieure d'Informatique, Oued-Smar, Alger, Algérie
{n_bousbia, s_ait_adda}@esi.dz

Résumé. Les systèmes d'apprentissage sont dédiés pour apprendre sur un domaine particulier organisé à travers des contenus numériques. Dans le cas de documents hypermédias, il est toujours utile de reconnaître le comportement de navigation de l'apprenant avant de pouvoir évaluer, personnaliser et adapter le processus d'apprentissage. Dans ce papier, nous visons l'analyse sémantique de la navigation de l'apprenant durant son apprentissage dans des contenus hypermédias. La raison principale de cette analyse est d'identifier le comportement de navigation d'un apprenant avec le cours étudié. Nous partons de l'hypothèse que si la distance sémantique entre les concepts du domaine des documents (ou pages) qui se suivent dans la navigation d'un apprenant est grande, alors ceci traduit un comportement déstructuré dans la navigation de l'apprenant et interprète une désorientation de ce dernier. Ce type de comportement pourrait être dû soit au faible niveau de connaissances de l'apprenant et/ou à une mauvaise organisation et structuration du cours. En effet, cette analyse va permettre au tuteur de reconnaître les apprenants désorientés et les aider, et à l'auteur du cours de restructurer et d'approfondir l'analyse des liens de navigation existants entre les parties du cours.

Mots-clés. Apprentissage en ligne, Evaluation du comportement de l'apprenant, Adaptabilité, Ontologie de domaine, Indexation conceptuelle.

1 Introduction

En tant que technologie d'apprentissage, les systèmes d'apprentissage sont en même temps utiles et problématiques. La virtualité, la flexibilité structurelle, l'interactivité... sont leurs atouts par rapport aux textes traditionnels, mais en même temps ils peuvent être aussi, la source de plusieurs problèmes, l'un des plus importants est la désorientation.

La désorientation est l'effet cognitif produit chez le lecteur qui ne fait plus le lien entre son projet de navigation initial et les zones d'informations qu'il est entrain de lire [1]. A l'origine, la désorientation est souvent due à la surcharge cognitive qui provient de l'incapacité de l'apprenant à retenir l'essentiel de l'information liée à un nœud parcouru, ou à ses difficultés d'identifier rapidement la nature des liens qu'il a activé [2]. L'apprenant doit se repérer, s'orienter et naviguer dans un espace d'informations en perpétuelle évolution. Cet embrouillement ne semble pas se ralentir

avec la croissance des réseaux. Face à ces obstacles (surcharge cognitive et désorientation) les concepteurs proposent des supports d'aide à la navigation [3]. Mais avant de pouvoir y remédier, il convient d'examiner un certain nombre d'aspects et d'y apporter un élément de précision. Il s'agit ainsi de se demander qui sont les apprenants en situation de désorientation et d'incohérence de lecture pour pouvoir les aider ?

Nous considérons en effet que répondre à cette question peut fournir des informations très significatives sur le déroulement de l'apprentissage, voire sur son efficacité. C'est ce que nous nous proposons d'exploiter dans cet article, à travers une méthode de modélisation implicite, basée sur le recueil et l'interprétation des traces d'interaction qui permet d'enrichir le modèle de l'apprenant en marquant le type de parcours. Nous proposons aussi d'introduire une ontologie du domaine enseigné afin de corréler les informations entre les différents contenus pédagogiques consultés par l'apprenant.

L'article sera organisé comme suit : dans la section 1, nous présentons certains travaux connexes qui ont pu soulever ce type de constat. En section 2, nous nous intéressons à certains outils du web sémantique, en particulier les ontologies. En section 3, nous développons l'approche proposée pour pouvoir détecter et évaluer le parcours d'un apprenant (désorienté ou pas). Nous terminerons avec des conclusions et des perspectives.

2 Etat de l'art

Plusieurs travaux ont essayé d'étudier la forme du parcours de l'apprenant pendant son apprentissage en étudiant la cohérence de lecture en rapport avec le niveau de l'apprenant.

Dans [4] les auteurs ont observé que les apprenants de faible niveau de connaissances faisaient davantage de sauts d'un sujet à l'autre alors que les apprenants de plus haut niveau de connaissances, poursuivaient des explorations plus complètes et profondes d'un même thème. Ford et Chen [5] ont observé des corrélations positives entre les connaissances antérieures et des comportements de navigation plus élaborés (moins de pages relues, exploration plus en profondeur, moins de demande d'aide) et la rapidité de la réalisation de la tâche d'apprentissage.

Dans une étude qualitative menée auprès de 8 participants, Rezende et de Souza Barros[6] ont identifié trois profils de navigation en fonction du niveau de connaissances antérieures : navigation « organisée » pour les apprenants de plus haut niveau de connaissances, navigation « conceptuelle » pour les apprenants de niveau intermédiaire et navigation « désorientée » pour les apprenants de faible niveau de connaissances. D'autres travaux, plus qualitatifs, vont dans le sens d'une navigation guidée par les connaissances en montrant qu'un manque de connaissances conduit les apprenants à poursuivre des stratégies peu élaborées (stratégies méthodiques et exhaustives), alors qu'un haut niveau de connaissances favorise l'emploi de stratégies moins erratiques [7].

Dans tous ces travaux étudiés, nous avons pu déduire qu'il est plus facile de reconnaître un apprenant à faible connaissance à travers la forme du parcours établi

pendant son apprentissage. En effet, notre travail s'inscrit sur cette optique, en essayant d'observer la forme de parcours de navigation d'un apprenant à travers l'application de mesure de similarité sémantique entre les concepts de pages qui se succèdent dans la navigation. Ceci dit, une distance sémantique assez grande entre les concepts de pages qui se succèdent dans la navigation traduit une incohérence de lecture de la part de l'apprenant et une errance de sa part (désorientation). Notons que les concepts du domaine enseigné sont bien modélisés à travers une ontologie du domaine propre au cours dispensé, comme nous le décrivons dans ce qui suit.

3 Ontologie du Domaine

Le web sémantique [8] est un espace compréhensible et navigable à la fois par l'être humain et les agents logiciels. Il introduit un complément de sens aux données de navigation du Web classique, basé sur les ontologies formelles et des vocabulaires contrôlés grâce aux relations sémantiques. Du point de vue de l'apprentissage en ligne, cela peut aider les apprenants à la localisation, l'accès, l'interrogation, le traitement et l'évaluation des ressources pédagogiques à travers un réseau hétérogène réparti, ou aider les enseignants dans la création, la localisation, ou encore le partage et l'échange d'objets d'apprentissage. Les technologies du web sémantique [9] ont également été utilisées comme une alternative pour permettre l'adaptabilité et l'interopérabilité des systèmes d'apprentissage.

Une ontologie [10] comprend un ensemble de termes, de connaissances, y compris le vocabulaire, des relations sémantiques, et un certain nombre de règles d'inférence logique associé aux concepts de l'ontologie, le tout relatif à un domaine particulier. L'ontologie appliquée au Web crée ainsi le Web sémantique [11].

L'utilisation des ontologies dans des environnements d'apprentissage tente de fournir des mécanismes visant à améliorer le processus de recherche et de découverte sémantique des ressources d'apprentissage. Elle offre par ailleurs la capacité d'organiser et d'afficher des informations telles que la visualisation des relations entre les concepts.

Le rôle de l'ontologie du domaine dans notre cas réside essentiellement dans l'indexation conceptuelle des pages consultés afin de faciliter leur identification. Dans notre cas d'étude, nous considérons que l'ontologie est composée d'un ensemble de concepts et relations entre ces concepts. Un identifiant unique est attribué à chaque concept, ces concepts sont labélisés avec un ou plusieurs terme synonyme. Expressément, une ontologie est définie comme suite : $O = \{C, R, Vo\}$.

- C : ensemble de concepts de domaine
- R : ensemble de relation entre les différents concepts du domaine
- Vo : le vocabulaire d'ontologie qui est composé de termes (simples ou composés) correspondant aux concepts de l'ontologie du domaine.

4 Approche Proposée

L'approche que nous proposons (cf. fig1) est basée sur l'analyse de fichiers de traces des apprenants en situation d'apprentissage lors de leurs navigations sur un cours portant sur un thème donné. Nous supposons que ce fichier contient essentiellement les Urls des pages visitées ainsi que leurs dates d'accès.

Ce fichier de traces va nous permettre de récupérer le parcours de navigation de l'apprenant, d'effectuer l'indexation sémantique des pages visitées et enfin de calculer l'indicateur de désorientation. Nous détaillons ces trois étapes en ce qui suit.

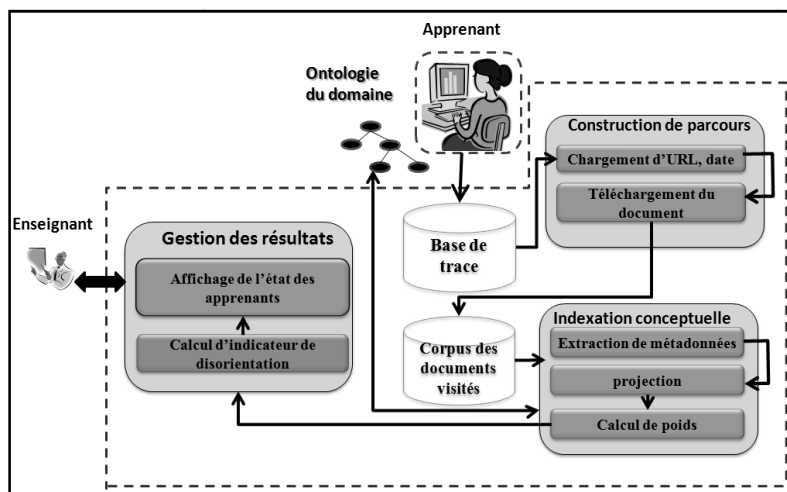


Fig1. Architecture de l'Approche proposée

4.1 Parcours de navigation

Il est évident qu'il est possible de représenter la navigation de l'apprenant comme un graphe de pages visitées, dans lequel les comportements des apprenants se traduisent par différents types de parcours.

En effet, nous pouvons représenter un parcours d'apprentissage pour une durée T par $G(V,E)$ tel que V représente l'ensemble des pages visitées et E est l'ensemble de liens activés par l'apprenant et reliant les différentes pages consultées pour son apprentissage

$$V = \{P_1, P_2, \dots, P_n\}, E = \{I_{1,1}, \dots, I_{n,m}\}.$$

Où $I_{i,j}$ indique un lien connectant la page P_i à la page P_j .

Donc après avoir eu les différentes pages parcourues par l'apprenant par le biais du fichier de traces, nous procédons à l'indexation des pages grâce à l'ontologie du domaine.

4.2 Indexation conceptuelle

Nous proposons d'utiliser une ontologie du domaine pour pouvoir indexer les documents parcourus, sachant que celle-ci est définie par le concepteur du cours. Le processus d'indexation est exécuté en quatre étapes principales : (1) représentation des documents consultés, (2) l'association des concepts aux mots-clés du document (3) désambiguïsation et (4) pondération des concepts.

a. Représentation des documents

Nous représentons chaque page visitée par ses mots-clés (grâce aux métadonnées). Les mots clés représentés par les métadonnées sont affectés aux documents d'une manière équitable, tandis que leur fréquence d'apparition, permet de donner importance d'un mot clé par rapport à un autre. Le modèle vectoriel est adapté dans l'approche proposée pour une représentation effective des documents. Chaque document est identifié par un vecteur de n dimensions, où chaque dimension correspond aux différents termes. A chaque terme du vecteur d'un document D_i est associé un poids :

$$D_i = (P_{1i}, P_{2i}, \dots, P_{nj}).$$

Où P_{ji} est le poids du $j^{\text{ème}}$ terme dans le document i .

Ce poids est une fonction de fréquence de terme, de collection et de facteur de normalisation, $tf*idf$ (termfrequency- inversed document Frequency)[12][13].

$$Tfidf_{i,j} = tf_{i,j} * idf_j \quad (1)$$

$$idf_j = \log(N/f_j) + 1 \quad (2)$$

Où $tf_{i,j}$ est la fréquence du terme j dans le document i , N est le nombre total de documents, f_j est le nombre de documents dans le corpus contenant le terme j .

Une fois que nous disposons des mots-clés de chaque document (le vecteur de termes) et leurs poids, nous procédons à leur projection sur l'ontologie du domaine.

b. Extraction des concepts

L'objectif de cette étape est d'identifier les concepts de l'ontologie qui correspondent aux mots-clés du document. L'extraction des concepts est basée sur l'analyse de chaque terme du document (mot-clé) avec toutes les entrées de l'ontologie [14]. Les concepts sont référencés dans les documents avec des mots-clés simples ou composés. L'algorithme d'identification est bien expliqué sur la figure 2.

Entrée: document D.
Sortie : VC, vecteur des concepts de l'ontologie représentant les mots-clés du document D.
Procédure
 Soit *mi* le prochain mot-clé à analyser dans le document D. Nous définissons le contexte **Par_i** par le paragraphe du document D qui inclut l'occurrence du mot *mi* à analyser.
Calculer $V_i = \{C_1, C_2, \dots, C_n\}$: les entrées de l'ontologie qui contiennent le mot *mi*.
Chaque $C_j \in V_i$ est représenté par un terme simple ou composé
Trier les concepts C_j de la collection V_i auquel : $|C(1)| > |C(2)| > \dots > |C(n)|$ // || représente la longueur du concept, en terme de nombre de mot correspondant au terme.
Pour chaque élément C_j dans V_i faire :
Obtenir le mot commun entre **Par_i** et le terme candidat de C_j , qui est l'intersection :
 $N = \cap (\text{Par}_i, C_j)$.
Si $|N| < |C_j|$ **alors** Le sens du concept C_j n'est pas inclus dans le contexte **Par_i**.
Sinon Si $|N| = |C_j|$ **alors** le sens du concept C_j est inclus dans le contexte **Par_i**.
 Ajouter alors C_j dans la collection des éléments du vecteur **VC**, associé déjà au document D (index).
Fin si
Fin pour

Fig2. Algorithme d'identification conceptuelle [15]

L'opération de conceptualisation des termes consiste alors à affecter à chaque terme *t* décrivant le document un concept *C* de l'ontologie du domaine. Néanmoins on risque de trouver un terme *ti* qui peut avoir *n* concepts possibles dans l'ontologie (plusieurs entrées dans l'ontologie). En conséquence, une étape de désambiguïsation est alors nécessaire.

c. Désambiguïsation

Dans un premier temps, on se sert de l'ontologie du domaine pour récupérer les différents sens possibles pour un terme ambiguë *ti* (concepts candidats). Puis, nous examinerons un terme non ambigu *tk*, dans le document, identifié par un concept *Ck* dans l'ontologie et auquel il est relié aux différents concepts candidats.

On calcule alors un score pour chacun des concepts candidats, ce score se base sur les mesures de similarités entre les différents sens du terme (concepts candidat) et le concept *Ck*, en utilisant des mesures tels que les mesures de Leacock et Chodorow (ou Lch) [16], de Lin [17] et de Resnik [18].

Dans notre cas d'étude, on utilise la méthode de Lin pour mesurer la similarité sémantique, cette méthode fait partie des plus étudiées sur le plan théorique et tient compte de l'information partagée par les deux concepts comme Resnik, mais aussi de ce qui les distingue :

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

Où $P(c)$: est la probabilité de rencontrer une instance du concept *c*.

$mscs(ci, ck)$: est le subsumant commun entre les deux concepts.

Le concept candidat *Ci* ayant le plus grand score est alors sélectionné pour représenter le terme *ti* ambigu.

d. Pondération de concepts

Les concepts ainsi extraits sont pondérés selon une méthode, plus générale que $tf*idf$, nommée Cf_c*idf (concept-frequency-inversed document frequency). Dans cette méthode chaque terme extrait représentera forcément un concept de l'ontologie étant donné qu'on a utilisé l'ontologie pour les identifier. Pour un concept C, sa fréquence dans un document dépend de la fréquence du terme lui-même [19]. Elle est calculée comme suit :

$$Cf_c = \sum_{tm \in t(c)} tf_{tm} \quad (4)$$

Où $t(c)$ est l'ensemble des différents termes correspondant au concept C. Le poids de chaque concept dans un document D est calculé comme suit :

$$Cfidf = Cf_c \times idf_c \quad (5)$$

idf_c est la fréquence inverse du concept C en comptant le nombre de documents dans lequel apparaît le concept C, analogiquement au poids de chaque terme dans un document D (cf. formule 2).

4.3 Calcul de l'indicateur de désorientation

Après l'indexation conceptuelle, chaque document (URL visitée) est représenté par un vecteur de concepts pondérés. Nous procédons alors au calcul de la distance sémantique entre deux vecteurs V1 et V2 correspondant à deux pages qui se suivent dans la navigation de l'apprenant.

Pour ce faire, nous avons opté pour l'indice de Jaccard [20]. La mesure de similarité de Jaccard est définie par le quotient du nombre des objets communs par le nombre total des objets auquel on soustrait le nombre d'objets communs :

$$sim(V1, V2) = \frac{Nc}{N1 + N2 - Nc} \quad (6)$$

Où N1 et N2 représentent les concepts des vecteurs V1 et V2 et Nc le nombre de concepts en commun entre les deux vecteurs.

Le résultat de la mesure de similarité est une valeur normalisée comprise dans l'intervalle [0,1].

L'indice de désorientation *des* correspond à la moyenne des distances de toutes les pages parcourues :

$$des = \frac{\sum_{i=0}^{n-1} sim(Vi, Vi + 1)}{n - 1} \quad (7)$$

Où n représente le nombre de pages parcourues par un apprenant

Le résultat de la mesure de désorientation *des* est encore une valeur normalisée comprise dans l'intervalle [0,1]. Par conséquent, tant que la mesure est grande et proche de 1, ceci indique un parcours de navigation cohérent, et dans le cas contraire (la mesure de *des* est proche du 0), ceci interprète un parcours incohérent, donc un apprenant en difficulté et qui nécessite de l'aide et du suivi. Un seuil α entre 0 et 1 sera fixé grâce à des expérimentations afin de préciser efficacement les apprenants en état de désorientation.

5 Conclusion

A travers notre étude nous avons décelé le besoin d'observer les comportements de navigation d'un apprenant afin de détecter le type de parcours qu'il entreprend pendant son apprentissage. Ce parcours est différent d'un apprenant à un autre selon le besoin et les difficultés rencontrées pendant l'apprentissage. La détection du type de parcours pourra ainsi nous aider à reconnaître les apprenants en difficultés, car un parcours déstructuré et incohérent est un parcours qui nécessite plus d'intention de la part du tuteur pour évaluer l'apprenant et lui apporter une éventuelle aide et de la part du concepteur du cours pour une révision et une restructuration du cours.

Pour cela, nous nous sommes intéressées essentiellement dans ce papier à l'indicateur de désorientation. Celui-ci est basé sur la distance sémantique entre chaque paire de pages qui se succèdent dans la navigation de l'apprentissage, or cet indicateur seul ne permet pas d'interpréter efficacement l'état et le comportement de l'apprenant. Nos perspectives se basent sur l'intégration de plusieurs autres indicateurs qui permettront de mieux évaluer et comprendre son état et comportement.

Références

1. Sahraoui, S.-A. : Problèmes d'intégration des Hypertextes/Hypermédias dans l'apprentissage des langues. Synergies Algérie. 1 (2007) 251-257
2. Amadiou, F., Tricot, A., Mariné, C.: Interaction between prior knowledge and concept-map structure on hypertext comprehension, coherence of reading orders and disorientation. *Interacting with computers*, 22(2), (2010)88-97
3. Rouet, J.F. : Interactivité et compatibilité cognitive dans les systèmes hypermédias», *Revue des Sciences de l'Éducation*(1999)
4. Carmel, E., Crawford, S., Chen, H.: Browsing in hypertext: a cognitive study". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22(5),(1992) 865-884.
5. Ford, N., Chen, S. Y. : Individual differences, hypermedia navigation, and learning: an empirical study. *Journal of educational multimedia and hypermedia*", Vol. 9(4), (2000)281-311.
6. Rezende, F., de Souza Barros, S.: Students' navigation patterns in the interaction with a mechanics hypermedia program. *Computers & Education*, 50(4), (2008)1370-1382
7. Mishra, P., Yadav, A.: Using hypermedia for learning complex concepts in chemistry: A qualitative study on the relationship between prior knowledge, beliefs, and motivation. *Education and Information Technologies*, 11(1), (2006)33-69
8. Dicheva, D. :*Ontologies and Semantic Web for E-Learning. Handbook on Information Technologies for Education and Training*, Springer Berlin Heidelberg, (2008)47-65

9. Siti, U., Rohiza, A., Shakirah, M. : Ontology of Programming Resources for Semantic Searching of Programming Related Materials on the Web. Information Technology (ITSim), 2010 International Symposium, IEEE, (2010) 698-703
10. Berners-Lee, T., Hendler, J., Lassila, O. : The Semantic Web. Scientific American, (2001) 29-37
11. Zschocke, T., deLeon, J. : Towards an Ontology for the Description of Learning Resources on Disaster Risk Reduction. CCIS 111, Springer-Verlag Berlin Heidelberg. WSKS (2010) 60-74
12. Robertson, S., Sparck Jones, K. : Relevance weighting for search terms. Journal of The American Society for Information Science, 27(3). (1976) 129-146
13. Sparck Jones, K. : Experiments in relevant weighting of search terms. Information Processing & Management, 15(3), (1979) 133-144
14. Baziz, M., Boughanem, M., Pasi, G., Prade, H. : 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) 301-313
15. Ait Adda, S., Balla, A. : 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) 106-114.
16. C. Leacock, G., Miller, A., Chodorow, M. : Using corpus statistics and WordNet relations for sense identification. Comput. Linguist. 24, 1 (Mar. 1998), 147-165.
17. Lin, D. An information-theoretic definition of similarity. In Proceedings of 15th International Conference On Machine Learning, (1998).
18. Resnik, P. 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), 11, 95-130. (1999).
19. Dragoni, M., Pereira, C. & Tettamanzi, A. : An Ontological Representation of Documents and Queries for Information Retrieval Systems, IEA/AIE 2010, Part II, LNAI 6097, Springer-Verlag Berlin Heidelberg, (2010) 555-564
20. Ljubešić, N., Boras, D., Bakarić, N., Njavro, J. : Comparing Measures of Semantic Similarity. Proceedings of the ITI 2008, 30th International Conference on Information Technology Interfaces, IEEE, (2008) 675 - 682