



HAL
open science

Indexation sémantique de ressources pédagogiques numériques. Approche par composants sémiotiques

Françoise Greffier, Federico Tajariol, Ioan Szilagyi

► **To cite this version:**

Françoise Greffier, Federico Tajariol, Ioan Szilagyi. Indexation sémantique de ressources pédagogiques numériques. Approche par composants sémiotiques. 7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2015), Jun 2015, Agadir, Maroc. pp.417-419. hal-01405978

HAL Id: hal-01405978

<https://hal.science/hal-01405978>

Submitted on 30 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Indexation sémantique de ressources pédagogiques numériques : une approche par composants sémiotiques

Françoise Greffier, Federico Tajariol, Ioan Szilagy
Objets et Usages Numérique (O.U.N.), Laboratoire ELLIADD, Montbéliard, France
Fédération de recherche FR Education, Besançon, France
Université de Franche-Comté
prenom.nom@univ-fcomte.fr

Résumé. Pour créer une ressource pédagogique numérique (RPN), un enseignant réutilise généralement les composants sémiotiques (e.g. un schéma, un texte) issus de ressources numériques existantes. Nous proposons d'instrumentaliser la tâche de sélection de composants à l'aide d'un parseur, qui repère et compte les composants. Pour guider l'enseignant-concepteur dans l'assemblage des composants sémiotiques, nous proposons un logiciel qui l'assiste dans la prise en compte des traitements cognitifs induits par ces composants.

Mots-clés. Ressource pédagogique numérique, composant sémiotique, ontologie

Keywords. Digital learning resource, semiotic component, ontology

1 Introduction

Une « ressource pédagogique numérique » (RPN) désigne un document numérique contenant du matériel pour l'enseignement. Toute RPN est le résultat d'une activité instrumentale menée par l'enseignant-concepteur (EC) via les TICE, qu'il s'agisse d'une création *ex nihilo* ou bien d'une ré-éditorialisation à partir de ressources numériques existantes [1]. Cette activité de ré-éditorialisation demande à l'EC un travail de *sélection* et de *recomposition*.

Pour la *sélection*, l'enseignant-concepteur doit ouvrir de nombreux fichiers, les lire pour repérer des fragments qui l'intéressent (e.g. un texte, une photo, un schéma, une animation, une bande son). Chaque fragment propose un contenu mis en forme et structuré : nous appelons ce fragment un « composant sémiotique » (CS) [2]. Nous considérons que cette phase de sélection serait plus aisée à l'aide d'un parseur qui indique les occurrences des types de composants sémiotiques d'une RPN et qui facilite leur localisation dans la ressource.

Pour la *recomposition*, l'enseignant-concepteur doit assembler les CS sélectionnés. Pour le guider dans cette phase complexe, nous proposons un logiciel qui assiste l'enseignant-concepteur à la prise en compte des traitements cognitifs

induits par les composants sémiotiques, car par exemple, la lecture d'un texte demande un traitement cognitif différent de la lecture d'un graphique [3]. Le but du logiciel est de constituer automatiquement des indicateurs sur les traitements cognitifs induits par les composants inclus dans une RPN.

2 Un parseur pour repérer les composants sémiotiques

Un composant sémiotique, dans un document numérique, est un élément textuel, imagé, auditif et audio-visuel [4]. Le parseur développé en Java, repère les types de composants d'une RPN donnée sous un format quelconque. Pour repérer les types de composants, le parseur extrait de la RPN des morceaux de texte et des images, à l'aide de bibliothèques. Or le type de composant texte ou image, n'est pas assez fin. Nous avons besoin de compter les morphèmes et de distinguer si une image présente un schéma ou une illustration. Nous avons conçu des algorithmes pour affiner les types en travaillant sur les matrices de pixels pour les images. Ainsi, des comparaisons sur le nombre de couleurs présentes dans la matrice de pixels nous permettent de distinguer une photographie d'un schéma. A ce stade du développement, le parseur repère et compte deux catégories de composants, les morphèmes et les images statiques (en noir et blanc ou en couleur) et leurs sous-catégories : icône, schéma, illustration [5]. Il reste des cas où il est difficile d'appliquer un algorithme pour définir le type d'un composant. Par exemple lorsque la RPN contient un lien hypertexte, le parseur peut repérer le lien, mais ne peut pas définir le type de CS pointé par le lien (e.g. une vidéo, une autre RPN). Des annotations sémantiques relatives à l'objet pointé pourraient nous permettre de lever cette incertitude.

3 Composants sémiotiques et traitements cognitifs

Nous proposons d'associer les composants sémiotiques aux habilités cognitives nécessaires pour traiter l'information suivant l'approche des Formes d'Intelligences Multiples [6]. Chaque forme d'intelligence comprend un système symbolique de codage de l'information. Ainsi nous avons pu établir une table de correspondances qui relie la mobilisation de certaines formes d'intelligence à la présence de type(s) de composants. Le démonstrateur que nous avons développé permet de calculer un pourcentage de « mobilisation » des formes d'intelligence sur une RPN au format *.ppt*. Il comprend 3 ontologies (développé avec Protégé version 5.0_beta, raisonneur Pellet) : une ontologie des composants sémiotiques, une ontologie des formes d'intelligence et une ontologie qualifiant la RPN. Les résultats du parseur sont introduits dans le démonstrateur par des assertions qui indiquent, pour chaque diapositive d'un fichier *.ppt*, le comptage des morphèmes et des images selon leurs sous-catégories. Le démonstrateur comprend des règles qui explicitent la table de correspondances et permettent au raisonneur de définir la mobilisation de formes d'intelligence sur chaque diapositive. Ensuite une requête SPARQL compte le nombre de fois où une forme d'intelligence est mobilisée par l'ensemble des diapositives. A ce stade du développement, le démonstrateur calcule les occurrences pour ces trois formes : linguistique, visuo-spatiale et logico-mathématique Les autres

formes d'intelligence ne sont pas encore repérables pour deux raisons. Soit parce ce que le parseur ne repère pas encore les composants associés (e.g. les sons). Soit parce que la liaison entre les composants et la forme d'intelligence va au-delà d'une simple présence. Par exemple, le repérage de la forme d'intelligence inter-personnelle exige une analyse des contenus de la RPN afin de déceler les formulations qui invitent, par exemple, à travailler en groupe. Un autre verrou que nous devons encore lever concerne la formule de calcul des pourcentages. Quel poids doit-on attribuer aux formes d'intelligence convoquées par une diapositive en fonction du nombre de CS rattachés à chacune de ces formes ? Par exemple, si l'on peut considérer que la présence d'un morphème suffit pour dire qu'une diapositive mobilise la forme d'intelligence linguistique, qu'en est-il d'une suite de morphèmes (e.g. une phrase) ?

4 Conclusion

Pour chaque RPN au format *.ppt*, notre système crée automatiquement les descripteurs sur le comptage et localisation des types de composants sémiotiques, ainsi que le pourcentage de formes d'intelligence associées. Ces descripteurs permettent l'indexation des RPN et la réutilisation des composants sémiotiques lors du processus de ré-éditorialisation. Notre système est développé sur la base d'ontologies afin de réaliser une indexation sémantique. Pour l'enseignant-concepteur, il est nécessaire de compléter le système par un logiciel de visualisation ergonomique de ces métadonnées. Nous préconisons de créer un logiciel de « feuilletage des RPN ». Celui-ci visualisera le nombre de composants par type ainsi que la liste des composants sémiotiques repérés et pour chacun une information sur sa localisation (e.g. numéro de diapositive). Jusque-là nous avons considéré les composants sémiotiques pris isolément. Nous envisageons par la suite, de prendre en compte les patterns de composants (e.g. un texte à gauche d'un schéma), pour repérer les plus fréquents, les localiser sur la RPN et pouvoir en décrire les traitements cognitifs associés.

Références

1. Crozat, S.: Chaînes éditoriales et rééditorialisation de contenus numériques. In L. Calderan, P. Laurent, H. Lowinger, & J. Millet (Éd.), *Le document numérique à l'heure du web* (2012) 179-220
2. Greffier, F., Tajariol, F : A semiotic and cognitive approach to the semantic indexation of digital learning resources In *Advances in Intelligent Systems and Computing. Lecture Notes in Computer Science* : Springer Verlag (in press).
3. Schnotz, W., Bannert, M.: Construction and interference in learning from multiple representation. *Learning and Instruction* 13 (2003) 141–156.
4. Moeschler J.: *Théorie pragmatique et pragmatique conversationnelle*. Paris, Armand Colin (1996).
5. Betrancourt, M.: The animation and interactivity principles in multimedia. In R. Meyer (ed.), *The Cambridge handbook of multimedia learning*. Cambridge, CUP, (2005) 287-296.
6. Gardner, H. : *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books. (2011).