

ANALYSE DE L'OBJET D'APPRENTISSAGE

1. Introduction

Notre approche d'assistance doit offrir aux auteurs les éléments nécessaires pour pouvoir prendre du recul vis-à-vis de l'objet conçu. A ce niveau de notre approche, l'auteur dispose déjà d'un objet d'apprentissage composé cohérent structurellement et sémantiquement. Au-delà de la cohérence, l'auteur doit vérifier que l'objet répond aux objectifs qu'il a fixé. De plus, il doit pouvoir vérifier que l'impact de la réutilisation d'autres objets composés n'a pas altéré la nature de l'objet par rapport ce qu'il a planifié au départ. L'analyse peut mener l'auteur même à découvrir certains aspects auxquels il n'a pas songé au départ. Ceci prend plus de sens dans le cas des auteurs non expérimentés et dans le cas des objets d'apprentissage complexes.

L'objectif, au niveau de cette phase, est donc de pouvoir fournir plusieurs mesures permettant à l'auteur de *comprendre* la vraie nature de l'objet conçu. Toutefois pour pouvoir effectuer des mesures il faut déjà définir des métriques. C'est dans cette optique que nous avons défini un certain nombre de métriques élémentaires [Defude, 2005] [Farhat, 2009a]. Ces métriques doivent être combinées et agrégées par la suite sous forme d'indicateurs de plus haut niveau et plus significatifs pour les auteurs des objets d'apprentissage. Dans la deuxième partie de ce chapitre nous allons présenter les notions d'espace sémantique et de composante sémantique. Nous allons examiner les différentes métriques sémantiques que nous pouvons en extraire. Nous allons également proposer diverses interprétations et utilisation de ces métriques. Les algorithmes utilisés pour l'extraction et pour le calcul de ces métriques sont également présentés.

2. Métriques

Les métriques élémentaires que nous avons définies peuvent être classées dans des ensembles. Le premier ensemble contient des métriques qui portent sur les propriétés structurelles.

2.1. Métriques Structurelles

Au niveau structurel nous avons des graphes orientés qui décrivent la logique de parcours de l'objet d'apprentissage composé. L'idée ici est de fournir à l'auteur des indications sur la complexité de la structure du graphe de composition. Un graphe est décrit par l'ensemble de ses sommets et par les arcs les reliant. Ceci nous permet de dégager deux métriques. La première métrique (M1) permettra de mesurer le nombre de sommets du graphe. La deuxième métrique (M2) permettra de mesurer le nombre des arcs dans le graphe.

Les sommets du graphe de composition peuvent être soit des objets d'apprentissage, soit des opérateurs ALT ou encore des opérateurs PAR. Ceci permet de dégager trois métriques supplémentaires. Une métrique (M3) permettra de mesurer le nombre de sommets de type objet d'apprentissage. Une métrique (M4) permettra de mesurer le nombre de sommets qui sont des opérateurs ALT. Une autre métrique (M5) permettra de mesurer le nombre de sommets qui sont des opérateurs PAR.

Voici la liste des métriques structurelles :

M1 : nombre de sommets

M2 : nombre d'arcs

M3 : nombre de sommets de type objet d'apprentissage

M4 : nombre des opérateurs de type ALT

M5 : nombre des opérateurs de type PAR

Ces métriques sont mesurables au niveau du graphe de composition abstrait et au niveau du graphe de composition concret.

2.2. Métriques Sémantiques

Plusieurs aspects sémantiques peuvent être mesurés via des métriques spécifiques. Une première métrique (M6) permet d'évaluer le nombre de concepts traités au niveau des contenus. Il s'agit d'un indicateur sur la densité sémantique de l'objet d'apprentissage.

Une deuxième métrique (M7) permet de mesurer le nombre de concepts requis au niveau des pré-requis de l'objet d'apprentissage composé. Il s'agit d'un indicateur sur la difficulté du contenu.

Nous avons également défini une métrique (M8) qui permet de mesurer le nombre de rôles éducatifs par concept. Il s'agit d'un indicateur d'ordre pédagogique.

La dernière métrique proposée (M9) est relative au nombre moyens de sommets par rôle éducatif. Il s'agit également d'un indicateur de la densité sémantique par rapport à chaque rôle.

Voici la liste des métriques relatives aux aspects sémantiques :

M6: *nombre de concepts au niveau du contenu*

M7: *nombre de concepts au niveau des pré-requis*

M8: *nombre moyen de rôles par concept*

M9: *nombre moyen de sommets par rôle*

Contrairement aux deux premières métriques (M6 et M7) les métriques M8 et M9 peuvent être utilisés pour faire des mesures d'une part sur le graphe de composition abstrait et d'autre part sur le graphe de composition concret.

2.3. Métadonnées

Les métriques proposées concernant les métadonnées ont pour objectif de donner des indicateurs à propos de la qualité de renseignement des métadonnées, ce qui a un impact direct sur l'accessibilité de l'objet d'apprentissage.

La première métrique (M10) permet de mesurer le nombre de catégories de métadonnées renseignées. Une catégorie est considérée comme renseignée si au moins une entrée (un champ) de cette catégorie est renseignée.

La deuxième métrique (M11) permet de mesurer le nombre total d'entrées de métadonnées renseignées. Il s'agit d'un indicateur supplémentaire sur l'accessibilité de l'objet d'apprentissage.

La troisième et la quatrième métrique (M12 et M13) permettent de mesurer le pourcentage respectivement de catégories et d'entrée de métadonnées renseignées. Il s'agit d'un indicateur sur la couverture des catégories et des métadonnées selon le profil d'application LOM retenu.

Voici un récapitulatif des métriques retenues :

M10: *nombre de catégories de métadonnées renseignées.*

M11: *nombre d'entrées de métadonnées renseignées.*

M12: pourcentage de catégories de métadonnées renseignées.

M13: pourcentage des entrées de métadonnées renseignées.

2.4. Niveaux d'abstraction

L'ensemble suivant de métriques porte sur des propriétés structurelles relatives aux différentes vues d'un graphe de composition d'un objet d'apprentissage. Nous les avons mis à part, par rapport aux métriques structurelles, vu que leur champ d'application est l'objet d'apprentissage et non pas chaque graphe de composition à part.

La première métrique (M14) correspond au niveau d'abstraction du graphe de composition abstrait. La mesure s'effectue en calculant le nombre de substitutions nécessaires pour convertir tous les objets composés en objets atomiques afin d'obtenir le graphe de composition concret.

La deuxième métrique (M15) permet de mesurer le nombre de graphes affichables. Il s'agit de calculer le nombre de parcours possibles au sein du graphe de composition concret.

La troisième métrique (M16) permet de mesurer le pourcentage de sommets présents dans tous les graphes affichables par rapport au total des sommets au sein du graphe concret.

Voici un récapitulatif des métriques d'abstraction :

M14: Nombre de niveaux d'abstraction.

M15: Nombre de graphes affichables.

M16: pourcentage de sommets du graphe concret présents dans tous les graphes affichables.

2.5. Synthèse

Rappelons que l'objectif de cette phase au sein de notre approche d'assistance est de fournir aux auteurs une cartographie détaillée des propriétés et caractéristiques de l'objet d'apprentissage conçu. Certes le concepteur de l'objet d'apprentissage est censé le mieux connaître les propriétés et les caractéristiques de son objet. Toutefois, certains aspects peuvent lui échapper surtout dès que l'objet atteint une certaine taille et/ou qu'il réutilise des objets créés par d'autres. Ainsi, réexaminer l'objet d'une façon assistée et automatisée ne peut être que bénéfique.

L'analyse doit être basée sur les valeurs qui correspondent aux différentes métriques qu'on a définies. Il faut noter que notre objectif ici est loin d'être l'exhaustivité ou la complétude,

mais de démontrer qu’il est possible de définir des métriques et d’effectuer des mesures sur les objets d’apprentissage. Ces mesures peuvent ensuite être combinées et agrégées pour produire des indicateurs de plus haut niveau comme nous allons le présenter dans la suite de ce chapitre. Le tableau suivant donne une synthèse des métriques et un aperçu sur la façon avec laquelle nous allons les utiliser pour générer suite à l’analyse des indicateurs de haut niveau aux auteurs des objets d’apprentissage composés :

	Métriques
Concordance structurelle	M1, M2, M3, M4, M5, M14
Capacité à être adapté	M4, M5, M15
Capacité à être réutilisé	M6, M7, M10, M11, M12, M13
Richesse didactique	M8, M9, M16

Il faut noter que d’autres métriques liées aux styles d’apprentissage d’un objet d’apprentissage ont été identifiés dans [Rojas 2008]. Dans ces travaux la classification de Honey et Mamford pour les styles d’apprentissage [Honey 1982], à savoir « *Activists* », « *Reflectors* », « *Theorists* » et « *Pragmatists* », a été retenue. Cette étude a donné lieu à quatre métriques relatives aux styles d’apprentissage. La première permet de mesurer le nombre de concepts théoriques. La deuxième permet de mesurer le nombre de vidéos, modèles visuels et images. La troisième permet de mesurer le nombre d’exercices pratiques. Alors que la dernière permet d’évaluer la possibilité de travailler en groupe.

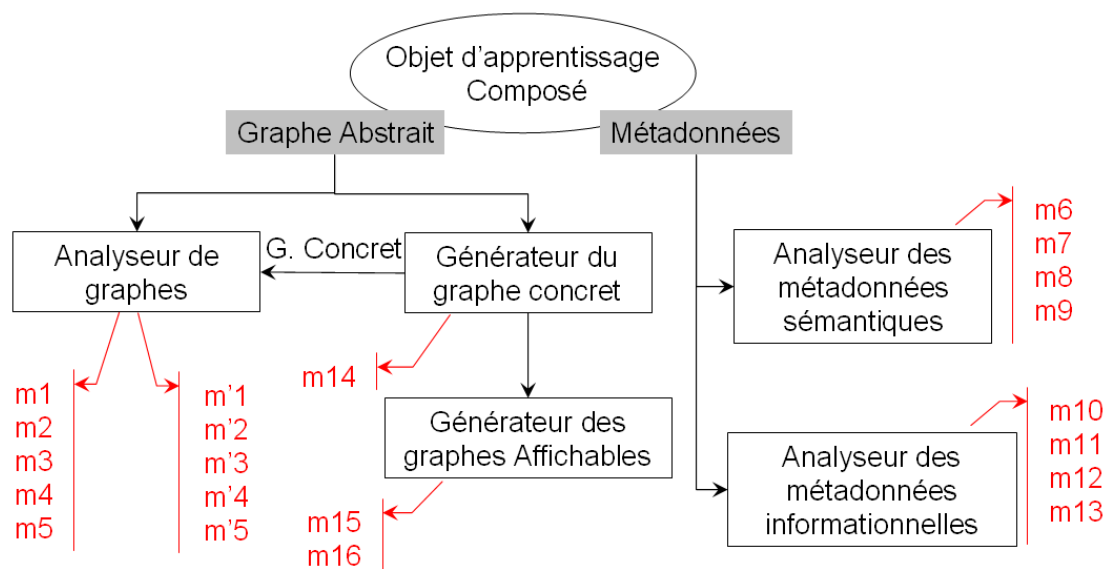
3. Indicateurs

Ces métriques de bas niveau seront difficilement utilisables par les auteurs. Pour leur donner un sens d’un point de vue métier il faut pouvoir les utiliser afin de produire des indicateurs de haut niveau compréhensible et significatif pour les auteurs d’objets d’apprentissage.

Les différentes métriques sont produites par différents modules. Trois des modules exploitent la description du graphe abstrait. En effet, le « générateur du graphe concret » permet en plus de sa fonction principale de générer la métrique relative aux niveaux d’abstraction (M14). Le

graphe de composition concret généré par ce module sera utilisé par le « générateur de graphes affichables » qui permettra de fournir les métriques relatives au nombre de graphes affichables (M15) et au pourcentage de sommets présents dans tous les graphes affichables (M16). Le graphe de composition concret et le graphe de composition abstrait seront analysés par « l'analyseur de graphes » afin de donner des mesures relatives aux métriques structurelles. Les métriques relatives au graphe de composition concret sont marquées avec une apostrophe.

Les métriques relatives à la sémantique et aux métadonnées sont générées par « l'analyseur de métadonnées sémantiques » et « l'analyseur des métadonnées informationnelles ».



Notation : m_i est la mesure associée à la métrique M_i

Figure 28 : Génération des métriques relatives aux objets d'apprentissage composés

Quatre indicateurs sont calculés à base des métriques relatives aux objets d'apprentissage composés : le degré de concordance structurelle, la capacité à être adapté, la capacité à être réutilisé et la consistance didactique. Dans un but d'uniformisation nous allons considérer une échelle de mesure qui varie de zéro à Max.

3.1. Concordance structurelle

Le premier indicateur est celui qui concerne la *concordance structurelle*. Pour cet indicateur une valeur proche de Max doit être attribuée s'il n'y a pas de différences entre la vue auteur (le graphe de composition abstrait) et la vue système (graphe de composition concret). Une valeur proche de zéro indiquera qu'une différence considérable est constatée entre la structure du graphe abstrait et celle du graphe concret.

$$V_1 = \left(c_1 \times \frac{m1}{m'1} + c_2 \times \frac{m2}{m'2} + c_3 \times \frac{m3}{m'3} + c_4 \times \frac{m4+1}{m'4+1} + c_5 \times \frac{m5+1}{m'5+1} + c_6 \times \frac{1}{m14+1} \right) \times \frac{Max}{\sum_{i=1}^6 c_i}$$

$V_1 \in [0, Max]$, en effet :

Pour les métriques m_i telles que $i \in [1,3]$, nous avons $\frac{m_i}{m'_i} \leq 1$ puisque la structure devient plus compliquée structurellement au niveau du graphe concret et donc toujours $m_i \leq m'_i$.

Pour les métrique m_4 et m_5 relatives au nombre des opérateurs ALT et PAR nous avons utilisé la formule $\frac{m_i + 1}{m'_i + 1} \leq 1$ pour éviter le cas de la division par zéro en cas d'absence d'opérateurs.

Pour la métrique m_{14} relative au niveau d'abstraction nous avons $\frac{1}{m_{14} + 1} \leq 1$ puisque d'une part m_{14} est inversement proportionnelle au degré de concordance structurelle et d'autre part cette valeur peut être nulle s'il n'y a pas d'objets composés dans le graphe de composition abstrait.

Chacun des termes de la formule est pondéré par un coefficient c_i qui permettra de paramétrer l'impact de chaque terme sur l'estimation du degré de concordance. La valeur finale est normalisée par la division de la somme des termes multipliés par leurs coefficients respectifs par la somme des coefficients. La valeur obtenue, qui est entre 0 et 1, est multipliée par la valeur maximum de l'échelle de mesure.

Rappelons que la valeur V_1 est égale à Max si le graphe concret et le graphe abstrait sont identiques (au cas où tous les sommets sont atomiques). Si la structure est plutôt la même au niveau du graphe abstrait et au niveau du graphe concret la valeur V_1 va avoir une valeur qui converge vers Max, ce qui indique une concordance structurelle élevée. Dans le cas contraire la valeur de V_1 va converger vers 0 ce qui reflète une faible concordance entre le graphe abstrait et le graphe concret.

C'est ce dernier cas qui devrait amener l'auteur à vouloir investiguer plus avant sur la cause de cette discordance structurelle et son impact sur l'objet d'apprentissage. Les valeurs des métriques utilisées peuvent donner des indications supplémentaires à l'auteur.

3.2. Capacité à être adapté

Le deuxième indicateur à déterminer concerne la capacité de l'objet d'apprentissage à être adapté aux profils des apprenants. Cette capacité dépend du nombre des opérateurs PAR et ALT, donc des métriques M4 et M5 du graphe concret. Ainsi lorsque leur nombre est élevé alors la capacité d'adaptation est importante. D'autre part cette capacité dépend du nombre de graphes affichables. Lorsque le nombre de graphes affichables est important alors la capacité de personnalisation est importante.

Ainsi, nous proposons la formule suivante :

$$V_2 = \left(c_1 \times \frac{m'_4 + m'_5}{m'_1} + c_2 \times \left(1 - \frac{1}{m_{15}} \right) \right) \times \frac{Max}{c_1 + c_2}$$

$V_2 \in [0, Max]$, en effet :

Le terme $\frac{m'_2 + m'_3}{m'_1}$ est inférieur à 1 puisque le nombre des opérateurs est forcément inférieur au nombre total des sommets du graphe de composition. La valeur du terme croît vers 1 lorsque le nombre des opérateurs PAR et ALT est important.

Le deuxième terme $1 - \frac{1}{m_{15}}$ varie de 0 vers 1 lorsque le nombre de graphes affichables est plus important.

Chacun des deux termes de la formule est pondéré par un coefficient c_i qui permettra de paramétrer l'impact de chaque terme sur l'estimation du degré de concordance. La valeur finale est normalisée par la division de la somme des termes multipliés par leurs coefficients respectifs par la somme des coefficients. La valeur obtenue, qui est entre 0 et 1, est multipliée par la valeur maximum de l'échelle de mesure.

3.3. Capacité à être réutilisé

Le deuxième indicateur concerne la capacité d'un objet d'apprentissage à être réutilisé. Cette capacité dépend des métadonnées et de la complexité sémantique de l'objet. Les mesures liées aux métadonnées sont utilisées dans des termes qui convergent vers 1 lorsqu'ils croissent et vers 0 si elles sont faibles. Quant aux termes qui utilisent les métriques liées au nombre des concepts dans le contenu et au nombre des pré-requis elles convergent vers 1 si les valeurs

sont faibles (pas beaucoup d'exigences au niveau des pré-requis et le contenu n'est pas trop chargé de concepts).

$$V_3 = \left(\frac{c_1}{m_6} + \frac{c_2}{1+m_7} + c_3 \times \left(1 - \frac{1}{m_{10}+1} \right) + c_4 \times \left(1 - \frac{1}{m_{11}+1} \right) + c_5 \times m_{12} + c_6 \times m_{13} \right) \times \frac{Max}{\sum_{i=1}^6 c_i}$$

$V_3 \in [0, Max]$, en effet :

Le terme $\frac{1}{m_6}$, pour la métrique M6 est inférieur à 1. Ce terme décroît dès que le nombre des concepts au niveau du contenu (respectivement au niveau des pré-requis) augmente, sachant que ce nombre ne peut pas être nul.

Le terme $\frac{1}{1+m_7}$, pour la métrique M7, est inférieur à 1. Ce terme décroît dès que le nombre des concepts au niveau des pré-requis augmente, sachant que ce nombre peut être nul et donc on l'additionne à 1.

Le terme $1 - \frac{1}{m_i + 1}$, pour les métriques M10 et M11, est inférieur à 1. En effet, au pire des cas, si la métrique a une valeur nulle le terme prend également la valeur nulle. Par contre, si les métadonnées sont totalement renseignées, la valeur du terme va tendre vers 1.

Les valeurs relatives aux métriques M12 et M13 sont des pourcentages. Ainsi, ils vont naturellement être compris entre 0 et 1.

Chacun des termes de la formule est pondéré par un coefficient c_i qui permettra de paramétrer l'impact de chaque terme sur l'estimation du degré de concordance. La valeur finale est normalisée par la division de la somme des termes multipliés par leurs coefficients respectifs par la somme des coefficients. La valeur obtenue, qui est entre 0 et 1, est multipliée par la valeur maximum de l'échelle de mesure.

3.4. Consistance didactique

Le quatrième indicateur porte sur la *consistance didactique* de l'objet d'apprentissage. Sa valeur croît lorsque le nombre moyen de rôles par concept, le nombre moyen de sommets par rôle et la similarité entre les graphes affichables sont importants.

$$V_4 = c_1 \times \left(\left(1 - \frac{1}{m_8} \right) + c_2 \times \left(1 - \frac{1}{m_9} \right) + c_3 \times m_{16} \right) \times \frac{Max}{\sum_{i=1}^3 c_i}$$

$V_4 \in [0, Max]$, en effet :

Le terme $1 - \frac{1}{m_i}$, pour les métriques M8 et M9, est inférieur à 1. En effet, les métriques relatives au nombre moyen de rôles par concept et au nombre moyen de sommets par rôle ne peuvent pas déjà être nulles. Si la valeur de la mesure augmente la valeur du terme va tendre vers 1.

La métrique M16 est un pourcentage. Ainsi, sa valeur est naturellement comprise en 0 et 1.

Chacun des termes de la formule est pondéré par un coefficient c_i qui permettra de paramétrer l'impact de chaque terme sur l'estimation du degré de concordance. La valeur finale est normalisée par la division de la somme des termes multipliés par leurs coefficients respectifs par la somme des coefficients. La valeur obtenue, qui est entre 0 et 1, est multipliée par la valeur maximum de l'échelle de mesure.

3.5. Exemple illustratif

Pour illustrer le processus d'analyse d'objet d'apprentissage composé nous allons considérer l'objet, intitulé lo_400, présenté par la figure suivante.

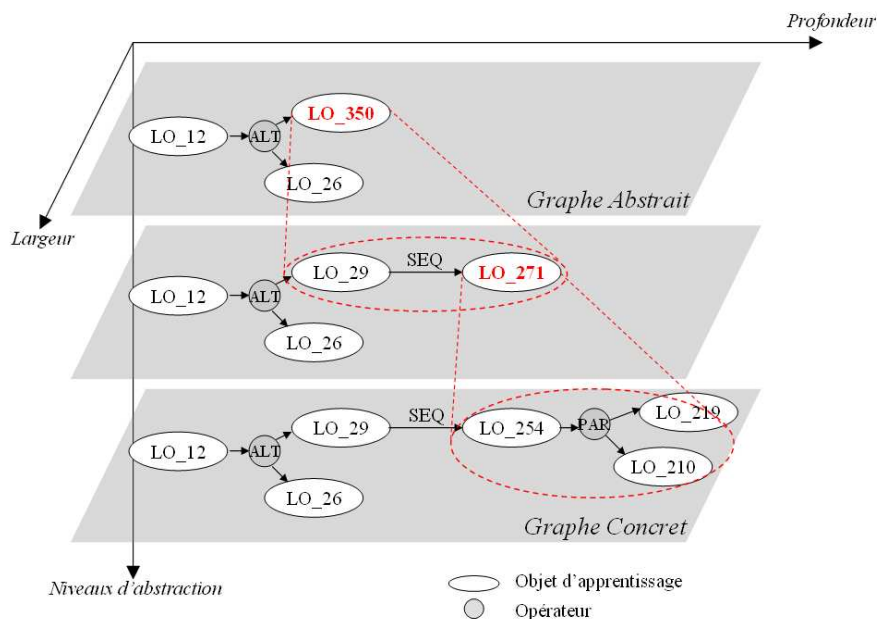


Figure 29 : Structure de l'objet d'apprentissage lo_400

Soient les métriques pour cet objet d'apprentissage composé :

Métriques liées à la structure

$m_1=4$; $m'_1=8$; $m_2=3$; $m'_2=7$; $m_3=3$; $m'_3=6$; $m_4=1$; $m'_4=1$; $m_5=0$; $m'_5=1$

Mesures relatives aux métriques liées à la sémantique

$m_6=2$; $m_7=1$; $m_8=3$; $m_9=2$

Métriques liées aux métadonnées

$m_{10}=3$; $m_{11}=10$; $m_{12}=0.33$; $m_{13}=0.17$

Métriques liées aux niveaux d'abstraction

$m_{14}=2$; $m_{15}=2$; $m_{16}=0.17$

Nous allons considérer une échelle de 10 (Max = 10) et des coefficients de pondération tous égaux à 1 ($c_i = 1$). L'application des différentes formules présentées auparavant permet d'obtenir les valeurs suivantes pour les divers indicateurs :

Degré de concordance structurelle	Capacité d'être adapté	Capacité d'être réutilisé	Richesse didactique
V1 = 5.4	V2 = 3.75	V3 = 6.1	V4 = 4.46

Ces résultats peuvent être présentés de diverses façons notamment graphiques :

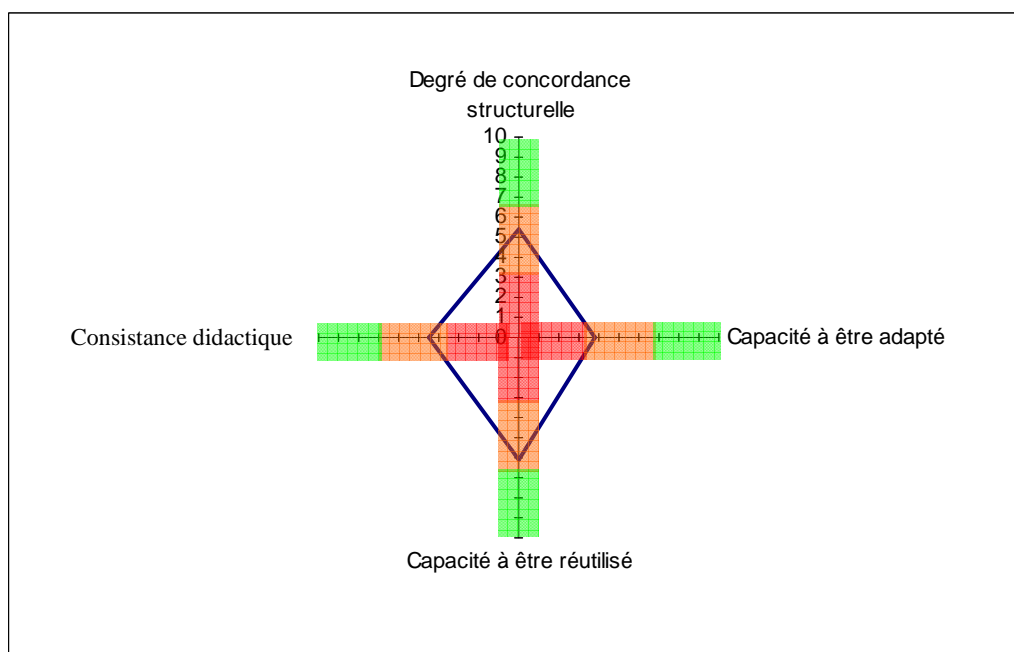


Figure 30 : Vue en mode radar des résultats d'analyse d'un objet d'apprentissage composé

Cette vue permettra à l'auteur d'avoir un résumé sur les caractéristiques de l'objet d'apprentissage qu'il a composé par assemblage d'objets existants. Dans le cas de notre exemple, l'auteur sera surtout alerté de la faible capacité de l'objet à être adapté. Ceci viens du fait qu'il y'en a que deux opérateurs chacun à l'origine de deux arcs.

4. Analyse de l'espace sémantique

Vu la nature didactique des objets d'apprentissage nous proposons un outil d'analyse supplémentaire dédié à la sémantique du contenu. L'objectif de cet outil est de permettre d'analyser le contenu de l'objet d'apprentissage en prenant comme référentiel le modèle de domaine. C'est dans ce contexte que nous introduisons la notion d'espace sémantique. En se basant sur cette notion, nous proposons des méthodes d'analyse permettant d'assister l'auteur afin de décortiquer les spécificités sémantiques de son objet d'apprentissage et de lui permettre d'améliorer son objet (en ajoutant ou en retirant des concepts du contenu).

Nous allons introduire au départ les notions d'espace sémantique et de composante sémantique telles que nous les avons définies. Ces notions seront utilisées afin de dégager les propriétés et caractéristiques relatives au contenu de l'objet d'apprentissage.

4.1. Notion d'espace sémantique

L'espace sémantique d'un objet d'apprentissage est défini comme étant l'ensemble des concepts traités au niveau du contenu d'un objet d'apprentissage, ainsi que les relations de type spécifique/générique qui les relient.

L'espace sémantique est déduit du modèle de domaine qui représente un référentiel commun dans les systèmes d'apprentissage en ligne sémantiques tel que SIMBAD. Pour y parvenir il suffit de marquer les concepts du modèle de domaine qui font partie du contenu et de retenir les relations spécifique/générique qui les relient. Ainsi, l'espace sémantique d'un objet d'apprentissage « clo » est défini de la façon suivante :

$$ES_{clo} = \{SC_{clo}, SR_{clo}\}$$

L'ensemble SC_{clo} est formé par les concepts du modèle de domaine qui décrivent le contenu de l'objet d'apprentissage « clo ». L'ensemble SR_{clo} est formé par les relations spécifique/générique qui existent entre les concepts qui font partie de SC_{clo} .

En fait, nous ne nous intéressons qu'aux relations de type spécifique/générique car nous ne voulons garder que la hiérarchie entre les concepts. Ce type de relations à notre avis est

suffisant pour pouvoir étudier les aspects cohérence, continuité et complexité de contenu d'un point de vue sémantique.

Soit un objet d'apprentissage, dont l'identifiant est LO_222, avec le contenu suivant :

```

{
  <Algorithme, Introduction>, <Algorithme, Définition>,
  <Algorithme, Exemple>, <Déclaration, Introduction>,
  <Type, Introduction>, <Type simple, Introduction>,
  <Variable, Définition>, <Variable, Exemples>,
  <Constante, Définition>, <Constante, Exemple>,
  <Corps, Introduction>, <Instruction, Définition>,
  <Ecriture, Définition>, <Ecriture, Exemple>,
  <Lecture, Définition>, <Lecture, Exemple>,
  <Affectation, Définition>, <Affectation, Exemple>,
  <Affectation, Exercice>, <Conditionnel, Définition>,
  <Conditionnel, Exemple>, <Conditionnel, Exercice>,
  <Itération, Définition>, <Itération, Exemple>,
  <Itération, Exercice>
}

```

Et soit cette version simplifiée d'un modèle de domaine qui couvre le domaine de l'algorithmique :

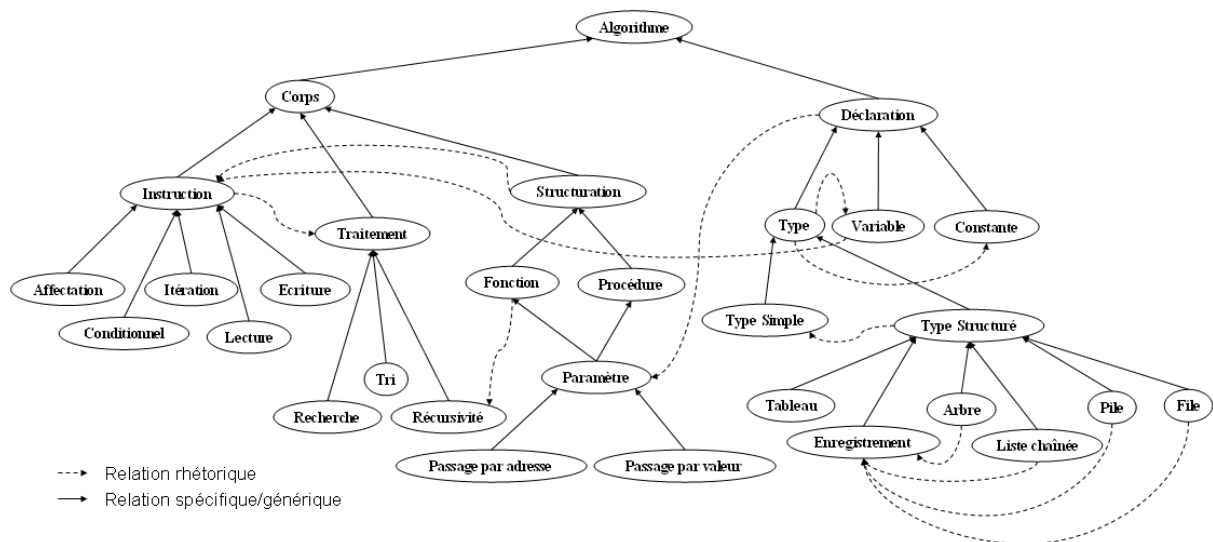


Figure 31 : Modèle de domaine (simplifié)

L'espace sémantique de l'objet d'apprentissage LO_222 est le suivant (c'est un sous-graphe du modèle de domaine) :

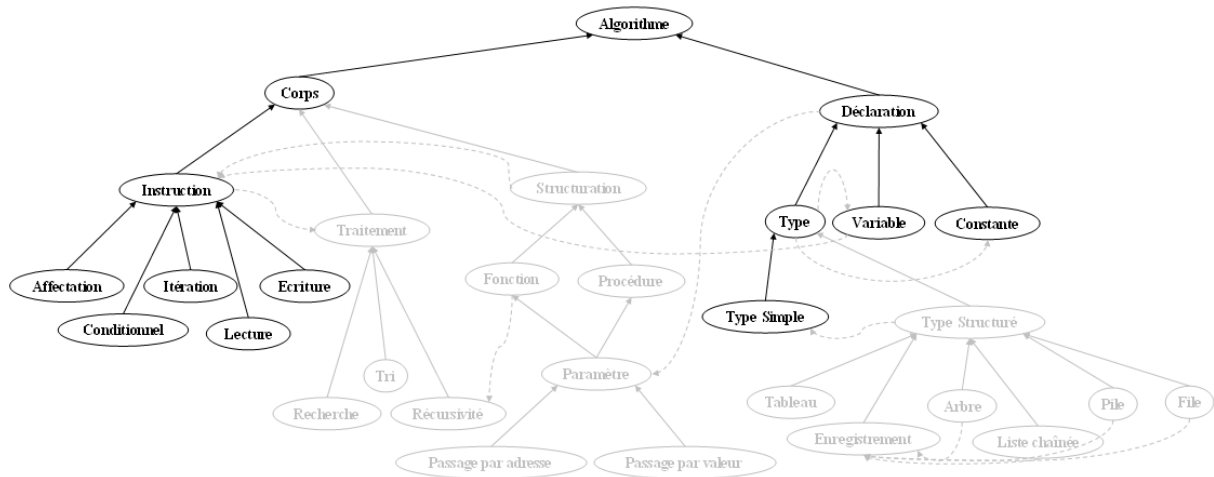


Figure 32 : Représentation graphique d'un espace sémantique

4.2. Calcul de l'espace sémantique

Nous allons noter O'_{clo} l'ensemble des objets d'apprentissage qui appartiennent au graphe de composition concret. Et soit R l'ensemble de toutes les associations de type spécifique/générique dans le modèle de domaine.

L'algorithme suivant permet de calculer l'espace sémantique d'un objet d'apprentissage composé :

Algorithme 9 : Calcul de l'espace sémantique

calculerEspaceSémantique (clo : ObjetComposé)

Debut

Pour tout ($i \in O'_{clo}$) **Faire**

Pour tout ($c \in C_i$) **Faire**

AddConceptToSC(c, SC)

Fin Pour

Fin Pour

Pour tout ($r \in R$) **Faire**

Si fromConcept (r, ci) et toConcept(r,cj) et $\{ci, cj\} \subset SC$ **Alors**

addRelationshipToSR (r, SR)

Fin

Fin

Fin

Afin d'automatiser l'interprétation de l'espace sémantique et d'en tirer les propriétés qui seront présentées à l'auteur de l'objet d'apprentissage, nous allons introduire le concept de composante sémantique.

4.3. Composante sémantique

Une composante sémantique est un groupement de concepts qui sont en proximité. La proximité sémantique se traduit au niveau de l'espace sémantique par une distance entre les concepts qui est inférieure à un seuil fixé. Cette distance correspond au chemin le plus court entre deux concepts donnés pour qu'ils puissent faire partie de la composante sémantique. Les composantes sémantiques forment des graphes connexes.

A titre d'exemple, la distance entre les concepts « Itération » et « Type simple » dans l'exemple ci-dessus est égale à cinq. D'autre part, la distance entre les concepts « Itération » et « Affectation » est deux. Ainsi, une distance faible reflète une proximité sémantique.

Ainsi, nous pouvons ainsi définir un espace sémantique d'un objet d'apprentissage comme un couple $\{CS, s\}$, avec CS est l'ensemble de composantes sémantiques et s est le seuil de proximité sémantique.

Nous pouvons définir une composante sémantique cs_i comme étant un ensemble de concepts C tel que : $\forall \{c_x, c_y\} \subset C, d(c_x, c_y) \leq s$

Prenons comme exemple l'objet d'apprentissage du paragraphe précédent. Définissant, l'espace sémantique comme étant l'ensemble de composantes sémantiques conformes à un seuil s égal à deux. La figure suivante montre les différentes composantes sémantiques au niveau de l'espace sémantique de l'objet d'apprentissage LO_222.

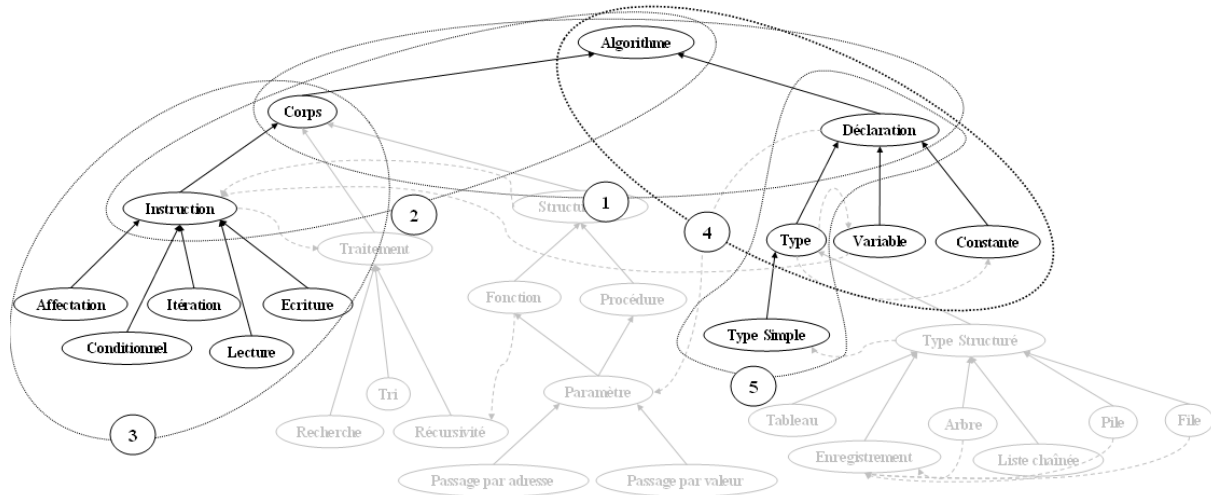


Figure 33 : Composantes sémantiques d'un espace sémantique

Nous allons utiliser ces notions d'espace sémantique et de composante sémantique afin d'analyser les caractéristiques sémantiques des objets d'apprentissage.

4.4. Complexité sémantique du contenu

L'un des facteurs qui peuvent rendre un contenu éducatif compliqué à assimiler est sa densité sémantique. Celle-ci peut être considérée comme étant proportionnelle au nombre de concepts traités au niveau du contenu. Sauf que le fait que certains concepts sont en proximité sémantique doit être pris en compte. En effet, si nous disposons de douze concepts traités au niveau d'un contenu éducatif et que ceux-ci font partie d'une même composante sémantique cela doit être plus simple à assimiler par l'apprenant que si ces concepts font partie de douze composantes sémantiques différentes. Rappelons qu'une composante sémantique reflète une proximité sémantique. Ce qui justifie le fait qu'acquérir douze concepts en proximité sémantique est certainement moins difficile qu'acquérir douze concepts qui font partie de douze composantes sémantiques. C'est-à-dire des concepts qui ne sont pas en proximité sémantique et donc nécessitent de la part de l'apprenant plus d'effort pour les assimiler. Ainsi, la complexité du contenu (d'un point de vue sémantique) peut être évaluée en fonction du nombre des concepts et du nombre de composantes sémantiques.

Soient x le nombre de concepts et y le nombre de composantes sémantiques au sein de l'espace sémantique d'un objet d'apprentissage.

Nous proposons la formule suivante pour évaluer la complexité sémantique du contenu :

$$\left(c_1 \frac{\log(x)}{\log(\text{Max}(\alpha, x))} + c_2 \frac{\log(y)}{\log(\text{Max}(\beta, y))} \right) \times \frac{E}{\sum c_i}$$

Les deux termes de la formule permettent d'avoir une variation logarithmique des valeurs permettant de contribuer à l'évaluation de la complexité du contenu. Si le contenu de l'objet ne traite qu'un seul concept alors forcément nous avons une seule composante sémantique. Les deux termes vont avoir la valeur minimale zéro car $\log(1) = 0$.

La valeur maximale de la complexité nécessite un avis d'experts qui vont décider du nombre de concepts α qui représente un seuil à partir duquel la complexité du contenu est considéré comme maximale. Un avis similaire est nécessaire pour avoir un seuil relatif au nombre de composantes sémantiques maximale β à partir duquel la complexité du contenu est considéré comme maximale. La nature du domaine de l'apprentissage, l'âge des apprenants, le niveau des apprenants et le niveau de détails retenu dans la description du modèle de domaine ont une influence sur les seuils α et β .

Finalement, pour normaliser la valeur maximale à un nous allons diviser la valeur logarithmique de x (respectivement de y) par le maximum entre α et x (respectivement de β et y).

Les coefficients c_1 et c_2 permettent de pondérer l'impact du nombre de concept et du nombre de composantes sémantiques sur la valeur estimative de la complexité sémantique du contenu. La valeur peut s'adapter à une échelle de mesures de 0 à E.

Dans le cas de l'objet LO_222 si on prend les coefficients c_1 et c_2 égaux à 1, une mesure sur une échelle de 10, une valeur α égale à 100 et une valeur β égale à 10, la complexité est calculée ainsi :

$$\left(\frac{\log(13)}{\log(100)} + \frac{\log(5)}{\log(10)} \right) \times \frac{10}{2} = 6,28$$

L'auteur peut en déduire que la complexité sémantique de cet objet d'apprentissage est plutôt moyenne.

Alors que s'il n'y avait que 2 concepts et une composante sémantique, par exemple, la complexité aura la valeur suivante : $\left(\frac{\log(2)}{\log(100)} + \frac{\log(1)}{\log(10)} \right) \times \frac{10}{2} = 0,15$

L'auteur peut en déduire que la complexité sémantique de cet objet d'apprentissage est faible.

D'autre part s'il y avait 50 concepts regroupés en 20 composantes sémantiques, par exemple,

$$\text{la complexité aura la valeur suivante : } \left(\frac{\log(50)}{\log(100)} + \frac{\log(20)}{\log(20)} \right) \times \frac{10}{2} = 9,25$$

L'auteur peut en déduire que la complexité sémantique de cet objet d'apprentissage est très élevée.

4.5. Continuité sémantique du contenu

Passer de l'étude d'un concept à l'étude d'un autre concept pour un apprenant est certainement plus aisé lorsque le passage se fait en faisant la transition via des concepts intermédiaires. Nous appelons cette propriété « continuité sémantique ». Cette facilité de passage d'un concept à un autre, que nous appelons continuité sémantique, est altérée au cas où les concepts sont éparpillés dans l'ontologie de domaine. L'espace sémantique de l'objet par conséquent contient des métriques qui permettent d'estimer le niveau de continuité sémantique d'un contenu.

En fait, nous disposons de deux métriques qui peuvent nous renseigner sur la continuité sémantique du contenu. La première métrique est le nombre de graphes connexes au sein de l'espace sémantique de l'objet d'apprentissage. En fait, si le contenu est parfaitement continu l'espace sémantique va former une seule composante connexe. Cependant, un nombre important de composantes connexes au sein de l'espace sémantique est le signe d'un contenu non continu sémantiquement.

La deuxième métrique est le nombre d'intersections entre les composantes sémantiques par rapport à leur nombre. Une intersection entre deux composantes sémantiques reflète l'existence d'au moins un concept en commun. Ce concept permet de faciliter le passage de l'étude des concepts faisant partie de la première composante sémantique à l'étude des concepts faisant partie de la deuxième composante sémantique.

Afin de produire un indicateur de la cohérence du contenu nous avons défini la formule suivante. Soit x le nombre de graphes connexes au niveau de l'espace sémantique et y le nombre d'intersections entre les n composantes sémantiques. La cohérence est calculée de la façon suivante :

$$\left(c_3 \frac{1}{(1 + \sum d_j)^x} + c_4 \frac{1}{1 + |\min(0, y - (n - 1))|} \right) \times \frac{E}{\sum c_i}$$

En fait, si le nombre de graphes connexes au niveau de l'espace sémantique est égal à un alors le terme $\frac{1}{x}$ prend sa valeur maximale qui est égale à un. Par contre si le nombre de composantes connexes est important alors la valeur de ce terme va tendre vers zéro. La quantité $\sum d_j$ correspond à la somme des distances entre les graphes connexes. La distance entre deux graphes connexes est la longueur du plus court chemin entre les deux graphes. Le coefficient c_3 est utilisé pour pondérer la contribution de ce terme dans l'évaluation de la continuité sémantique de l'objet d'apprentissage.

Le deuxième terme utilisé est $\frac{1}{1 + |\min(0; y - (n - 1))|}$ où n est le nombre de composantes

sémantiques et y est le nombre d'intersections entre elles. En fait, pour n composantes sémantiques le nombre d'intersection minimal pour un contenu qui a une continuité sémantique est égale à $n-1$. Donc, si y est égal à $n-1$ ou plus le terme doit prendre une valeur égale à un. Si le nombre y est strictement inférieur à $n-1$ alors certainement il y en a une discontinuité sémantique. Ainsi, nous allons utiliser la fonction minimum pour avoir zéro dans le cas où la valeur de y est supérieure ou égale à $n-1$ et donc le terme prend une valeur égale à 1. Si y est strictement inférieur à $n-1$ alors $y-(n-1)$ aura une valeur négative qui sera retournée par la fonction minimum et donc le terme va avoir une valeur qui tend vers 0 si la différence entre le nombre d'intersections et le nombre de composantes sémantiques est importante. La valeur absolue de la valeur retournée par la fonction minimum permet d'avoir une valeur positive pour le terme. Le coefficient c_4 est utilisé pour pondérer la contribution de ce terme dans l'évaluation de la continuité sémantique de l'objet d'apprentissage.

Ainsi, les coefficients c_3 et c_4 permettent de pondérer l'impact du nombre de graphes connexes entre les composantes sémantiques d'une part, et le nombre d'intersections entre les composantes sémantique d'autre part sur la valeur estimative de la continuité sémantique du contenu. La valeur peut s'adapter à une échelle de mesures de 0 à E.

Dans le cas de l'objet LO_222 et avec des coefficients c_3 et c_4 égaux à un et une mesure sur une échelle de dix, la continuité est calculée comme suit :

$$\left(\frac{1}{1} + \frac{1}{1 + |\min(0; 7 - (5 - 1))|} \right) \times \frac{10}{2} = 10$$

L'auteur peut en déduire que son objet d'apprentissage satisfait pleinement la propriété de continuité sémantique.

Par contre si l'objet comporte quatre composantes sémantiques avec une seule intersection. Et si l'espace sémantique de l'objet comporte trois graphes connexes dont la somme des distances qui les sépare est 7, alors nous avons la valeur suivante :

$$\left(\frac{1}{(1+7) \cdot 3} + \frac{1}{1 + |\min(0; 1 - (4-1))|} \right) \times \frac{10}{2} = 1,87$$

Cette valeur reflète une continuité sémantique faible qui mérite d'être examinée par l'auteur.

4.6. Déséquilibre sémantique du contenu

L'espace sémantique d'un objet d'apprentissage décrit les différents concepts traités par cet objet. Cependant, il ne permet pas de savoir si les concepts sont traités d'une façon équilibrée par les objets d'apprentissage réutilisés (au niveau du graphe de composition abstrait) ou si certains concepts ou certaines composantes ont été traités plus que d'autres.

Pour pouvoir étudier cet aspect nous proposons d'étiqueter chaque concept de l'espace sémantique par l'objet d'apprentissage qui le traite. Le nombre d'objets qui traitent un concept représente son « poids » sémantique. De la même façon nous pouvons également déterminer le poids de chaque composante sémantique.

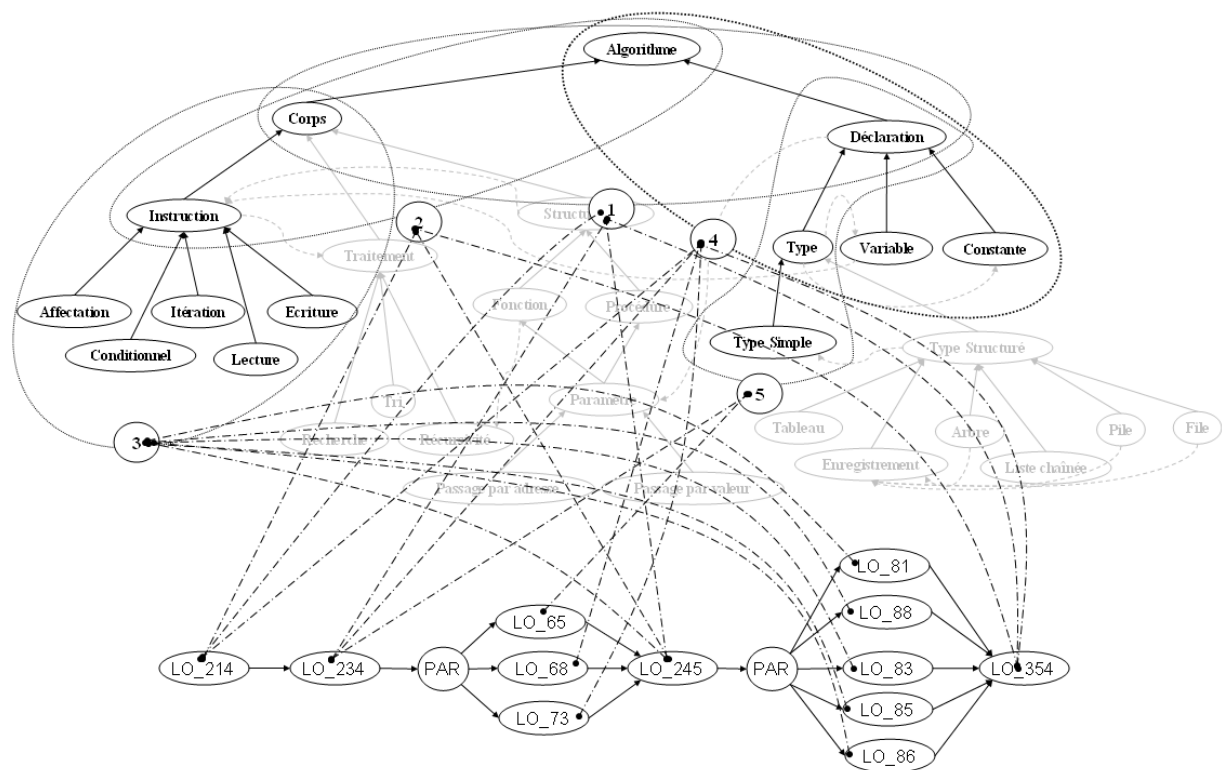


Figure 34 : Etiquetage des composantes sémantiques par les objets d'apprentissage

Cette information doit permettre à l'auteur de l'objet d'apprentissage composé de réexaminer son objet dans le cas où un grand déséquilibre est constaté. Si nécessaire l'auteur peut soit réduire le nombre de concepts traités en retenant les concepts qui ont un poids important et en éliminant ceux qui sont considérés comme marginaux. Sinon, l'auteur peut décider d'ajouter d'autres objets d'apprentissage à son graphe pour rétablir un certain équilibre entre les poids des concepts et des composantes sémantiques.

	LO_214	LO_234	LO_65	LO_68	LO_73	LO_245	LO_81	LO_88	LO_83	LO_85	LO86	LO_354	Tot
1	X	X				X						X	4
2	X					X						X	3
3						X	X	X	X	X	X		6
4	X	X		X	X							X	5
5		X	X										2

Tableau 1 : Poids des composantes sémantiques de l'objet d'apprentissage LO_222

En fait, nous disposons de deux métriques qui peuvent nous renseigner sur le déséquilibre sémantique éventuel du contenu. La première métrique est la différence entre le poids maximal et le poids minimal des différents concepts au niveau de l'espace sémantique.

La deuxième métrique est la différence entre le poids maximal et le poids minimal des différentes composantes sémantiques au niveau de l'espace sémantique de l'objet d'apprentissage.

Afin de produire une estimation du déséquilibre sémantique d'un objet d'apprentissage nous proposons la formule suivante. Soient M le poids maximal et m le poids minimal parmi les poids des concepts au sein de l'espace sémantique. Et soient M' le poids maximal et m' le poids minimal parmi les poids des composantes sémantiques. Soit D la différence maximale possible entre les poids. Cette valeur correspond au nombre d'objets d'apprentissage au sein du graphe de composition abstrait (poids maximal possible) diminuée par un (poids minimal possible). L'estimation est calculée de la façon suivante :

$$\left(c_5 \frac{\log(1 + (M - m))}{\log(1 + D)} + c_6 \frac{\log(1 + (M' - m'))}{\log(1 + D)} \right) \times \frac{E}{\sum c_i}$$

En fait, les deux termes de la formule opèrent de la même façon. S'il y a un grand écart entre le poids minimal et le poids maximal alors la valeur du terme doit tendre vers un. S'il y en a un équilibre sémantique parfait, alors le poids minimal est égal à zéro.

Les coefficients c_5 et c_6 permettent de pondérer la contribution des termes dans l'évaluation de l'équilibre sémantique. La valeur estimative prend une valeur sur une échelle de 0 à E.

Dans le cas de l'objet LO_222 et avec des coefficients c_5 et c_6 égaux à un et une mesure sur une échelle de dix, le déséquilibre sémantique est calculée comme suit :

$$\left(\frac{\log(1 + (3 - 1))}{\log(1 + 11)} + \frac{\log(1 + (6 - 2))}{\log(1 + 11)} \right) \times \frac{10}{2} = 5,45$$

L'auteur peut en déduire que l'objet présente un déséquilibre sémantique moyen qui n'est pas forcément inquiétant.

Toutefois, si l'objet avait la valeur douze comme poids maximal au niveau des concepts, et un comme poids minimal. Et si l'objet avait douze comme poids maximal au niveau des composantes sémantiques et un comme poids minimal, nous avons un déséquilibre maximal qui est révélé par la valeur estimative :

$$\left(\frac{\log(1 + (12 - 1))}{\log(1 + 11)} + \frac{\log(1 + (12 - 1))}{\log(1 + 11)} \right) \times \frac{10}{2} = 10$$

En revanche, si l'équilibre sémantique est idéal entre les concepts et entre les composantes sémantiques alors les valeurs $(M-m)$ et $(M'-m')$ sont égaux à zéro. A ce moment nous n'avons pas de déséquilibre et la valeur estimative donne zéro :

$$\left(\frac{\log(1 + 0)}{\log(1 + 11)} + \frac{\log(1 + 0)}{\log(1 + 11)} \right) \times \frac{10}{2} = 0$$

5. Apports de l'analyse

Nous considérons la phase d'analyse comme un élément clé dans notre approche. Le fait que l'auteur dispose déjà d'un objet d'apprentissage cohérent structurellement et sémantiquement lui procure la sensation que le processus de conception est terminé. Toutefois, pour rendre l'approche effectivement itérative et incrémentale il faut pousser l'auteur à examiner de nouveau son objet en lui facilitant cette tâche.

En effet, nous nous sommes mis comme objectif de fournir d'une façon automatisée des indicateurs divers et variés à propos d'un objet d'apprentissage composé. Ceci étant un moyen

pour permettre à l'auteur de l'objet de pouvoir identifier les éventuelles carences, lacunes et/ou faiblesses au niveau de l'objet par rapport aux objectifs qu'il s'est fixé au départ.

Nous avons démontré que les métriques de bas niveau que nous avons définies et calculées peuvent être combinées et agrégées pour offrir une information de plus haut niveau et orientée métier sous forme d'indicateurs. Toutefois, ceci a été fait à titre indicatif et ne constitue pas le cœur de notre contribution.

En effet, ces métriques constituent une matière première qui doit permettre aux acteurs de l'ingénierie pédagogique de pouvoir étudier la meilleure façon de les interpréter. Ces métriques constituent des éléments de réflexion pour le développement d'approches, méthodes et outils spécifiques permettant d'interpréter ces informations de la façon la plus efficace. Notons en plus que la façon avec laquelle ces informations seront présentées aux auteurs doit être étudiée. Ceci relève de l'ingénierie des interfaces homme/machine.

Un système implémentant notre approche d'assistance pourra exploiter cette analyse de diverses façons. Par exemple, le résultat de l'analyse peut être positionné par rapport à des valeurs de référence. Ces valeurs pourront être les valeurs moyennes de tous les objets présents dans une banque d'objets d'apprentissage, les valeurs moyennes des objets les plus utilisés par les utilisateurs ou encore les valeurs moyennes des objets déjà composés par le même auteur.

Cette phase d'analyse peut être également exploitée au-delà d'un environnement auteur classique. En effet, nous avons étudié son intégration dans une banque d'objets d'apprentissage. Ceci nous a permis de l'intégrer à titre expérimental dans la banque d'objets d'apprentissage COLORS (COoperative Learning Object Repository System) [Farhat et Jemni, 2009] qui sera décrite dans la partie implémentation.

Dans le système COLORS nous proposons à l'auteur suite à l'importation de son objet au niveau de la banque de le soumettre à une phase d'analyse. Les mesures effectués sont ensuite comparées à celles qui correspondent à ceux des dix objets d'apprentissage les plus consultés par les utilisateurs de la banque.

6. Conclusion

Dans ce chapitre nous avons mis en évidence dans un premier temps la possibilité d'extraire des informations implicites à partir des caractéristiques des objets d'apprentissage. Nous avons défini dans ce cadre des métriques qui couvrent les aspects structurels, sémantiques,

métadonnées et niveaux d'abstraction. En se basant sur ces métriques nous avons défini les algorithmes nécessaires pour générer les mesures adéquates.

Nous avons proposé une interprétation de ces valeurs calculées en définissant des fonctions d'agrégation. Ce travail a permis de donner des estimations sur la concordance structurelle de l'objet d'apprentissage composé, sur sa capacité à être adapté, sur sa capacité à être réutilisé et sur sa richesse d'un point de vue didactique.

Nous avons également proposé les notions d'espace sémantique et de composante sémantique. En se basant sur ces notions nous avons défini des formules paramétrables permettant de donner des valeurs estimatives purement relatives aux caractéristiques sémantiques de l'objet d'apprentissage composé sujet de l'analyse. En effet, nous avons pu donner des estimations significatives pour le niveau de complexité, pour la cohérence et pour le déséquilibre sémantiques d'un objet d'apprentissage.