



Intégration des connaissances de domaine dans un processus d'ECD multi-vues

El Moukthtar Zemmouri, Hicham Behja, Abdelaziz Marzak, Brigitte Trousse

► To cite this version:

El Moukthtar Zemmouri, Hicham Behja, Abdelaziz Marzak, Brigitte Trousse. Intégration des connaissances de domaine dans un processus d'ECD multi-vues. 4èmes Journées Francophones sur les Ontologies, Jun 2011, Montréal, Canada. hal-00940784

HAL Id: hal-00940784

<https://hal.archives-ouvertes.fr/hal-00940784>

Submitted on 2 Feb 2014

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.

Intégration des connaissances de domaine dans un processus d'ECD multi-vues

EL Moukhtar Zemmouri* — **Hicham Behja*,*****
Abdelaziz Marzak** — **Brigitte Trousse*****

* *Laboratoire CCSP*

*Ecole Nationale Supérieure d'Arts et Métiers - Meknes
Marjane 2 B.P. 4024 Beni M'hammed Meknes Maroc*

ezemouri@yahoo.fr, h_behja@yahoo.com

** *Laboratoire TIM*

Faculté des Sciences Ben M'sik, Casablanca, Maroc

marzak@hotmail.com

*** *INRIA Sophia Antipolis - Méditerranée, Equipe-projet AxIS*

2004 route des Lucioles BP 93 06902 Sophia Antipolis France

brigitte.trousse@inria.fr

RÉSUMÉ. Le processus d'Extraction de Connaissances à partir de Données (ECD) est un processus itératif et interactif complexe faisant appel à plusieurs types de connaissances et de savoir faire. Dans ce papier nous proposons d'appréhender la complexité d'un processus d'ECD multi-vues, c.-à-d. un processus mené par plusieurs experts avec différents points de vue. Notre objectif est d'assurer une meilleur coordination et compréhension entre les différents analystes d'une analyse multi-vues. Ainsi, nous proposons une modélisation de la notion de point de vue en ECD en se basant sur le standard CRISP-DM, et en tenant compte des connaissances expertes de domaine mises en œuvre lors d'une analyse d'ECD.

ABSTRACT. Knowledge Discovery in Databases (KDD) is a highly complex, iterative and interactive process involving several types of knowledge and expertise. In this paper we propose to support users of a multi-view analysis (a KDD process held by several expert with different viewpoints). Our objective is to enhance coordination and comprehensibility between experts. To do so, we propose a formalization of viewpoint in KDD based on CRISP-DM standard and taking into account domain knowledge involved during a KDD process.

MOTS-CLÉS : Connaissances de domaine, point de vue, processus d'ECD, ontologie, CRISP-DM, OWL.

KEYWORDS: Domain knowledge, viewpoint, KDD process, ontology, CRISP-DM, OWL.

1. Introduction

Le processus d'Extraction de Connaissances à partir de Données est un processus itératif et interactif complexe ayant pour objectif « d'identifier, dans des données, des patterns ultimement compréhensibles, valides, nouveaux et potentiellement utiles » (Fayyad et al., 1996). Il s'effectue sur trois étapes principales (prétraitement de données, fouille de données et post-traitement des patterns extraits) interrompues continuellement par des prises de décisions de l'utilisateur expert (figure 1).

Les connaissances de domaine jouent un rôle fondamental durant toutes les phases du processus d'ECD. En effet, un utilisateur du processus est confronté à plusieurs défis. D'une part il doit se familiariser avec les données à analyser, traduire son objectif d'étude en tâche d'ECD et sélectionner les attributs pertinents pour l'étude. Ce qui fait appel aux connaissances sur le domaine analysé. D'autre part il doit être capable de bien choisir, paramétrer, composer et exécuter des outils et méthodes provenant de divers domaines (statistique, intelligence artificielle, bases de données...) afin de réussir son objectif d'analyse. Ce qui fait appel aux connaissances du domaine de l'analyste (figure 1).

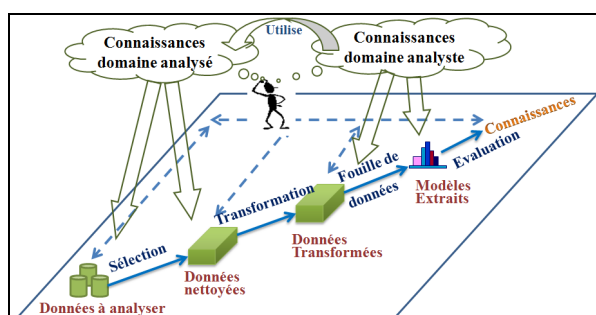


Figure 1. Interaction entre connaissances de domaine et étapes du processus d'ECD (Behja et al., 2010)

De plus, un processus d'ECD est le plus souvent mené par un ou plusieurs experts, chacun ayant ses préférences, son domaine de compétence, ses objectifs et sa propre vision des données, des méthodes et des fonctions d'ECD. Ainsi le processus d'ECD sera guidé par le point de vue de l'analyste (Behja et al., 2010) et plusieurs types de connaissances et de savoir-faire sont mis en jeu.

La figure 2 montre un exemple d'une analyse multi-vues des données d'une plateforme d'enseignement à distance. Ces données peuvent être analysées par plusieurs analystes acteurs du système : enseignants, tuteurs, administrateur, marketing, etc. L'objectif de l'analyse de l'enseignant (ex. évaluation du taux d'apprentissage d'un cours) n'est pas le même que celui de l'administrateur (ex. étudier la fiabilité du système), les attributs utilisés pour l'évaluation d'un cours sont

différents de ceux utilisés pour étudier la fiabilité de la plateforme. De même, les méthodes, outils et techniques d'ECD utilisés seront différents, et l'interprétation des résultats du processus dépend des objectifs d'analyse, et de la vision de l'analyste. Ainsi, il est fondamental de tenir compte du point de vue de chaque analyste et d'intégrer les deux types de connaissances de domaine dans le processus d'ECD.

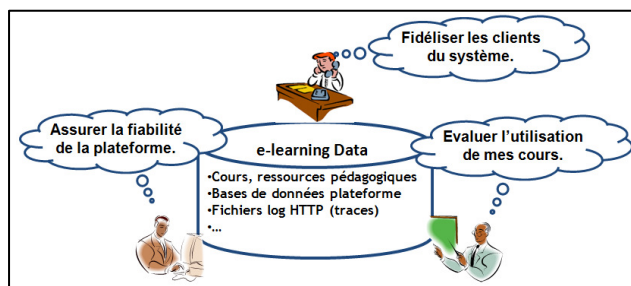


Figure 2. Multitude des points de vue d'analyse des données d'une plateforme e-learning

Plusieurs travaux ont traité de la complexité du processus d'ECD avec différentes approches pour supporter les utilisateurs (novices ou experts) d'ECD. La plupart de ces approches proposent à l'utilisateur d'explorer l'espace des plans d'exécution valides (Bernstein et al., 2005)(Diamantini et al., 2009), la découverte et l'accès aux services distribués de data mining (Euler, 2005)(Diamantini et al., 2009), la réutilisation des expériences de data mining réussies (Morik et al., 2004), etc. Sans prendre en compte l'aspect multi-vues d'une analyse d'ECD.

Nous proposons dans ce papier de mieux appréhender la complexité d'une analyse d'ECD multi-vues. Notre objectif est d'assurer une bonne coordination et compréhension entre les différents analystes. Pour ce faire, nous proposons une modélisation de la notion de point de vue en ECD en suivant une démarche d'ingénierie des connaissances : identification, modélisation, et opérationnalisation (Charlet et al., 2000). Dans la phase d'identification nous allons nous baser sur le standard CRISP-DM¹ pour identifier un ensemble de critères génériques qui permettent de caractériser un point de vue en ECD. Dans la deuxième phase, les connaissances mises en œuvre lors d'une analyse multi-vues seront structurées dans un modèle conceptuel de connaissance composé de quatre sous-modèles hiérarchiques : modèle du domaine, modèle de tâches et méthode, modèle du point de vue et modèle organisationnel des points de vue. Dans la troisième phase, le modèle du point de vue sera formalisé en utilisant le langage OWL² (Web Ontology Language).

1. CRISP-DM project, <http://www.crisp-dm.org/>

2. OWL, W3C Consortium, <http://www.w3.org/TR/owl-ref/>

Après cette introduction où nous avons cadré la problématique de notre travail, le reste du papier est organisé comme suit : dans la deuxième section nous présentons notre approche de point de vue en ECD et sa situation par rapport à d'autres approches en ingénierie des connaissances. Dans la troisième section, nous présentons les critères génériques permettant d'instancier un point de vue en ECD. La quatrième section présente notre modèle conceptuel de gestion des connaissances dans un processus d'ECD multi-vues ainsi que la formalisation du modèle de point de vue. Enfin des conclusions et perspectives dans la section 5.

2. La notion de point de vue en ECD

La notion de point de vue est une notion polysémique introduite en ingénierie des connaissances depuis les années 70 (Minsky, 1975), surtout pour la modélisation et le développement de systèmes complexes qui sont par nature multi-vues. Depuis, plusieurs travaux ont porté sur les significations, représentations, interprétations et confrontation des points de vue, avec parfois des appellations différentes : perspective, opinion, contexte, vue, etc.

L'approche multi-vues a été utilisée dans différents domaines : dans les méthodes et langages orientés objets VBOOM (View Based Object Oriented Method), VBOOL (View Based Object-Oriented Language) ; dans la programmation par vue; dans les bases de données objets; dans la modélisation du processus de développement; et aussi dans le langage UML. Une étude détaillée et une comparaison de ces approches peuvent être consultées dans (Nassar, 2004).

Le concept de point de vue a été aussi abordé implicitement ou explicitement par une large communauté de l'ingénierie des connaissances. Notamment dans la représentation des connaissances par objet, KRL et son amélioration LOOPS, ROME et son extension pour le langage de Frames FROME, TROPES; dans l'élaboration des bases de connaissances multi-expertes C-VISTA, dans la construction d'ontologies multi-points de vue MVP et en représentation des connaissances en conception de systèmes complexes. Une étude détaillée et une comparaison de ces approches peuvent être consultées dans (Bach, 2006).

Selon les travaux et les domaines cités ci-dessus, la définition du mot point de vue varie. Ribière (Ribière et al., 2002) propose une classification des définitions du concept point de vue en deux catégories qui correspondent à deux interprétations possibles : perspective et opinion. Les *points de vue perspectives* font référence aux différentes positions conceptuelles des experts par rapport à un objet (système, problème, connaissances...). Ce qui permet « d'indexer des descriptions consensuelles d'un même objet par différents experts », et aussi d'accéder à un sous ensemble d'informations ou connaissances pertinentes selon un point de vue donnée. Les *points de vue opinions* quant à eux font référence à des avis donnés par des experts sur un objet. La notion d'opinion est souvent en relation étroite avec l'expert et elle tient compte de son expérience, ses connaissances, ses tâches, etc.

Nous proposons de rendre explicite la notion de point de vue en ECD. En effet un processus d'ECD est un processus complexe (système complexe) le plus souvent mené par plusieurs experts (donc multi-vues) (voir exemple figure 2). La définition que nous proposons pour la notion de point de vue en ECD se place principalement dans une approche ingénierie des connaissances (connaissances mises en œuvre lors d'une analyse d'ECD) et prend en considération les deux facettes perspective /opinion définies par Ribière. Notre objectif est de modéliser le point de vue en ECD et de capturer la sémantique d'un processus d'ECD en termes de point de vue. Ainsi, nous proposons la définition suivante :

Définition : *Un point de vue en ECD est une interface permettant (1) l'accès à un sous-ensemble de connaissances de domaine (domaine analysé et domaine de l'analyste) et conduisant l'analyste à atteindre son objectif d'analyse, et (2) de capturer la logique du raisonnement et la trace des décisions prises par l'analyste lors d'une analyse d'ECD (i.e. capturer la sémantique du processus).*

Ainsi le point de vue de l'analyste lui permet de filtrer les connaissances expertes pertinentes selon sa vision sur (1) les données analysées, (2) le domaine d'application, et (3) le domaine d'ECD (tâches, méthodes, fonctions et outils d'ECD), et aussi selon son objectif de l'étude.

Lors de la spécification de son point de vue, l'analyste va s'appuyer sur l'instanciation de critères génériques que l'on peut décliner selon trois composantes: domaine analysé, domaine de l'analyste et contexte de l'analyse (figure 3).

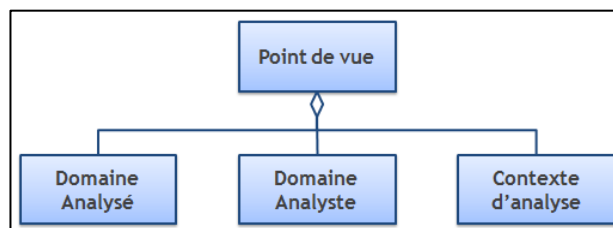


Figure 3. Composantes d'un point de vue en ECD

Pour identifier ces trois composantes du point de vue, nous nous sommes inspirés des points de vue de la cybernétique de second ordre : objet observé, observateur et contexte d'observation (Ben Ahmed, 2005). Il s'agit d'une vision systémique du processus d'ECD multi-vues qu'on analyse selon trois points de vue : point de vue domaine analysé (l'objet observé), point de vue domaine de l'analyste (l'observateur), et point de vue contexte d'analyse.

Pour identifier les critères génériques caractérisant un point de vue en ECD, nous nous sommes basés sur le standard CRISP-DM. Ce point fait l'objet de la section suivante.

3. Identification des critères générique d'un point de vue en ECD

Notre objectif à présent est d'identifier l'ensemble des critères génériques qui caractérisent un point de vue en ECD. Une fois instanciés, ces critères définissent le point de vue d'un analyste. Un point de vue qui va orienter l'exécution du processus d'ECD, et permettre par la suite de garder la trace du raisonnement effectué pendant l'analyse.

Les critères du point de vue sont appelés génériques « s'ils ne sont pas instanciés » (Rivière et al., 2002) (par exemple *KDD_Task* est un critère générique), ils sont appelés critères spécifiques s'ils sont instanciés (par exemple *KDD_Task = "Description"* est un critère spécifique). Les critères du point de vue sont génériques aussi dans le sens où ils sont indépendants du domaine d'application et des outils de data mining. Pour cette raison, nous avons jugé utile l'élicitation de ces critères génériques en se basant sur le standard CRISP-DM. La généricité des critères sera assurée par le niveau d'abstraction et de description du processus selon CRISP-DM.

En effet, CRISP-DM (CROSS-Industry Standard Process for Data Mining) est un modèle de processus et une méthodologie qui décrit les approches couramment utilisées par les experts pour une bonne conduite de projets décisionnels. Le modèle CRISP-DM est générique, indépendant du domaine d'application des techniques et outils de data mining. La méthodologie CRISP-DM est décrite sous forme d'un ensemble de tâches structurées en quatre niveaux d'abstraction : phases, tâches génériques, tâches spécifiques et instances de processus (du plus générique au plus spécifique). Au niveau le plus haut, un processus d'ECD selon CRISP-DM est typiquement organisé en six phases (figure 4) :

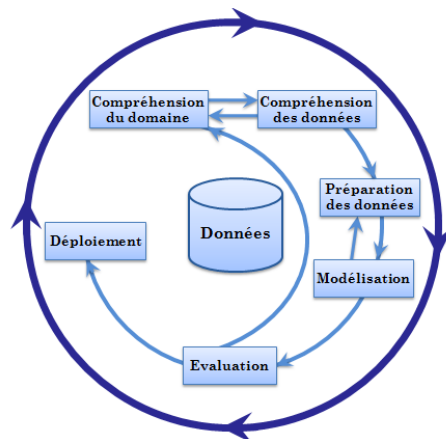


Figure 4. Phase du processus d'ECD selon le modèle CRISP-DM.

Dans chacune des phases du processus, l'analyste doit exécuter un certain nombre de tâches et préparer des livrables. Selon les connaissances requises pour chaque tâche, on distingue : Les tâches qui font appel aux connaissances du domaine analysé ; Les tâches qui font appel aux connaissances du domaine de l'analyste, son savoir faire et ses compétences ; et les tâches qui font appels aux objectifs poursuivis et aux critères de validation des résultats du processus (les modèles). Cette classification des tâches nous a permis de dresser le tableau des critères suivant :

	Domaine analysé	Domaine de l'analyste	Contexte d'analyse
Compréhension du domaine	<ul style="list-style-type: none"> • Background • Inventaire des ressources • Exigences, hypothèses et contraintes du projet • Risques et incertitudes • Terminologie 	<ul style="list-style-type: none"> • Tâche d'ECD • Critères de réussite de la tâche d'ECD • Plan d'exécution initial • Evaluation initiale des outils et des techniques 	<ul style="list-style-type: none"> • Objectifs du projet (objectifs métiers) • Critères de succès du projet • Coût et bénéfices du projet
Compréhension Des données	<ul style="list-style-type: none"> • Description collecte de données : sources, méthodes, problèmes • Description des données : format, quantité • Qualité des données : incomplètes, manquantes, erreurs 	<ul style="list-style-type: none"> • Exploration des données 	
Préparation Des données	<ul style="list-style-type: none"> • Données sélectionnées • Données nettoyées • Données transformées • Données préparées • Description des données 	<ul style="list-style-type: none"> • Méthodes de sélection et paramétrage • Méthodes de nettoyage et paramétrage • Méthodes de prétraitement et paramétrage 	
Fouille Des données	<ul style="list-style-type: none"> • Hypothèses sur les données • Données d'apprentissage, données de teste validation 	<ul style="list-style-type: none"> • Techniques de fouille à utiliser • Procédure de test validation • Paramètres d'exécution des méthodes et outils • Modèles résultats • Description des modèles • Evaluation des modèles 	
Evaluation		<ul style="list-style-type: none"> • Modèles approuvés • Révision du processus • Listes des actions possibles • décisions 	<ul style="list-style-type: none"> • Évaluation des résultats du processus d'ECD par rapport aux critères de succès initiaux
Déploiement		<ul style="list-style-type: none"> • Plan/stratégie de déploiement • Plan de maintenance • Révision et expérimentation des résultats du processus d'ECD 	<ul style="list-style-type: none"> • Rapport et présentation finals des résultats du processus d'ECD

Tableau 1. Attributs d'un point de vue en ECD selon le modèle CRISP-DM.

4. Modèle conceptuel de gestion des connaissances

Selon le projet CommonKADS, un modèle conceptuel en ingénierie des connaissances « est une spécification des informations et des structures et fonctions des connaissances mise en œuvre dans une tâche à forte implication des connaissances » (Schreiber et al., 2000). Il permet une description abstraite des connaissances indépendamment de l'implémentation. On y distingue trois types de connaissances qui font l'objet de modèles distincts avec des primitives de modélisation propres : les connaissances du domaine, les tâches et les méthodes (Charlet et al, 2000).

Notre modèle conceptuel de gestion des connaissances intégrant la notion de point de vue (figure 5) est composé de quatre sous-modèles structurés selon (Aussenac-Gilles, 1996) en connaissances du domaine et connaissances stratégiques. Le niveau domaine décrit les concepts du domaine et leurs relations. Le niveau stratégique se base sur le niveau domaine et exprime comment une tâche va être effectuée.

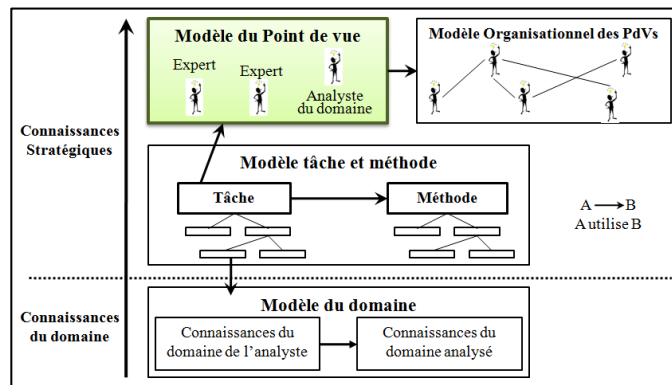


Figure 5. *Modèle conceptuel de gestion des connaissances dans un processus d'ECD multi-vues.*

Dans ce papier nous allons nous focaliser beaucoup plus sur la description et l'opérationnalisation du modèle de point de vue de la figure 5.

4.1. *Modèle du domaine*

Les connaissances du domaine sont les connaissances relatives au domaine d'application et nécessaires pour que les méthodes puissent s'exécuter. Le modèle du domaine est une conceptualisation des concepts et relations du domaine d'application. Dans notre contexte d'ECD, on considère ce modèle comme une ontologie de domaine utilisée pour indexer les données et les attributs manipulés.

4.2. Modèle tâche et méthode

Ce modèle décrit de façon abstraite le processus d'extraction en termes de tâches et de méthodes. Les tâches sont réalisées par des méthodes. Une tâche est une description de ce qui doit être fait dans l'application en termes de buts et de sous-buts. Les méthodes décrivent comment un but peut être atteint en termes d'une série d'opérations et d'un ordre de réalisation.

Nous avons formalisé ce modèle sous forme d'une ontologie générique semi-formelle *OntoECD* (Zemmouri et al., 2009). Cette ontologie conceptualise les méthodes et fonctions du processus d'ECD indépendamment du domaine d'application et de la structure des données, ce qui favorise la généricité et la réutilisabilité. *OntoECD* devrait répondre aux questions relatives au processus d'ECD, telles que les différents buts d'un processus d'ECD, les étapes, les méthodes implémentant ces étapes, les sous méthodes de bas niveau qui, une fois exécutées, réalisent la tâche de la méthode mère, ainsi que les caractéristiques des outils logiciels impliqués dans l'analyse.

4.3. Modèle du point de vue

Le modèle du point de vue est une conceptualisation des critères génériques introduits au paragraphe 3. Ces critères sont indépendants de la tâche et du domaine d'application. Ils permettent de modéliser la vision de l'analyste sur les données analysées, l'objectif d'analyse et une partie de l'expertise nécessaire aux nombreuses prises de décisions effectuées pendant l'analyse. La modélisation du point de vue en ECD devrait favoriser la coordination et la compréhension entre les différents experts d'une analyse multi-vues. Elle favorise également la réutilisation d'une analyse selon un point de vue donné (ceci est possible grâce à une annotation du processus que nous avons introduit dans Behja et al. 2010).

Pour formaliser le modèle du point de vue, nous avons opté pour le langage OWL (Web Ontology Language) vu son expressivité par rapport à XML schéma et RDF(S), et vu ses capacités tant représentationnelles qu'inférentielles. La figure 6 présente une partie de la hiérarchie des classes et relations définissant un point de vue en ECD.

Un point de vue en ECD est composé de (figure 6) : une composante domaine de l'analyste (*Analyst_Domain*) qui décrit les actions et décisions prises par l'analyste en termes de tâche d'ECD (vérification, description, prédiction), méthodes exécutées, et critères de validation des méthodes. Une composante domaine analysé (*Analyzed_Domain*) décrit la vision de l'analyste sur le domaine analysé en termes de données sélectionnées, attributs pertinents pour l'analyse, ontologie de domaine d'application, et effet de l'exécution des méthodes sur les données. Une composante contexte d'analyse (*Context*) décrit l'objectif décisionnel du projet ainsi que les critères de validation des résultats (modèles).

La composante domaine de l'analyste est principalement une description des actions et des décisions prises par l'analyste pour la construction d'un plan d'exécution (KDD_Execution_Plan). Un plan d'exécution (figure 6) est une composition d'un certain nombre de méthodes (préparation des données et modélisation selon CRISP-DM). Chaque méthode doit être choisie et configurée en fonction de la tâche et des données en main.

La composante domaine analysé est principalement une description des données manipulées et de l'effet de l'exécution des méthodes sur ces données. Les données (figure 6) ont un certain nombre de propriétés : format, source, qualité (valeurs manquantes, données erronées...), et quantité (nombre d'attributs, volume de données...). Chaque méthode exécutée par l'analyste a un effet transformationnel sur le jeu de données et ses propriétés.

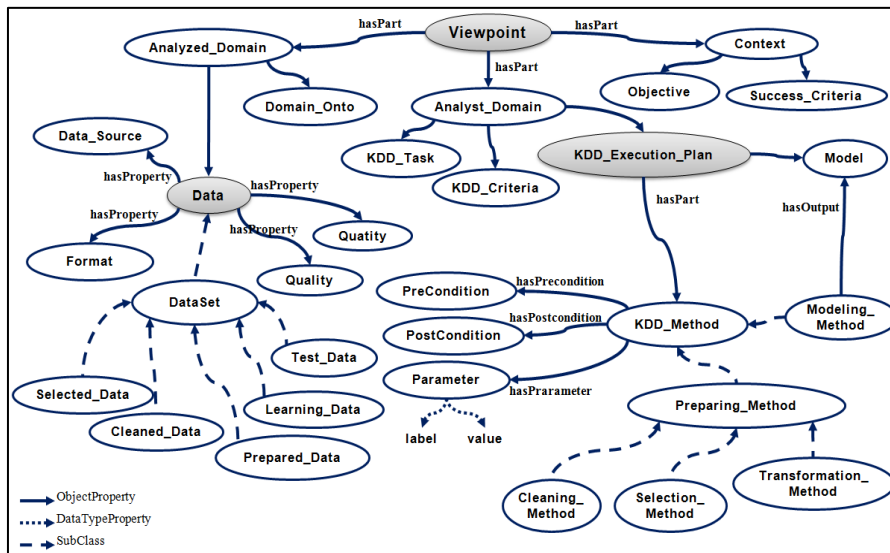


Figure 6. Hiérarchie des classes définissant un point de vue en ECD.

L'instanciation des classes caractérisant le point de vue en ECD permet de définir le point de vue d'un analyste. Ce point de vue permet de guider l'exécution du processus d'ECD et de garder la trace des décisions prise par l'analyste.

4.4. Modèle organisationnel des points de vue (figure 5)

Une analyse multi-vues est caractérisée par l'intervention de plusieurs expert (ex. figure 2). Il est donc important de modéliser l'interaction et la dépendance entre les différentes analyses selon des points de vue différents. Le modèle organisationnel des points de vue nécessite l'identification de divers types de relations entre points

de vue comme: équivalence, exclusion, complémentarité, composition, subsumption. Ce modèle est en cours d'élaboration.

5. Conclusions et perspectives

Dans ce papier nous avons présenté une modélisation de la notion de point de vue en ECD qui intègre les deux types de connaissances mises en œuvre dans un processus d'ECD : connaissances du domaine analysé et connaissances du domaine de l'analyste. Cette approche multi-vues du processus d'ECD devrait favoriser la coordination et la compréhension entre les différents utilisateurs d'une analyse multi-vues, et également la réutilisation d'une analyse selon un point de vue donné. Par exemple, il devrait y avoir complémentarité entre le point de vue « évaluation » (figure 2) d'un enseignant et le point de vue « fiabilité » de l'administrateur, l'enseignant peut consulter les résultats de l'administrateur pour mieux comprendre et interpréter ses résultats d'analyse. Un enseignant peut aussi réutiliser une analyse réussie d'un autre enseignant pour évaluer ses cours.

La figure 9 présente la version actuelle d'une plate-forme que nous développons pour supporter des analyses multi-vues. Elle permet la définition d'un point de vue par instanciation de quelques uns des critères génériques définis dans ce papier (contexte, tâche, sélection de données, attributs pertinents), puis de guider l'analyste dans le choix, le paramétrage et l'exécution des méthodes d'ECD en se basant sur la bibliothèque d'algorithmes Weka³. Le choix des méthodes se base surtout sur la nature des données sélectionnées par l'analyste et sur les pré- et post-conditions de chaque méthode. Mais, d'une manière générale, ce sont les attributs du point de vue de l'analyste qui devront permettre de choisir les méthodes convenables.

Le système a une architecture web services (Zemmouri et al., 2010) et utilise l'API Jena⁴ et son moteur de requêtes SPARQL⁵.

Nous projetons de développer d'avantage le modèle organisationnel des points de vue qui permet de définir des relations entre points de vue, et d'une manière générale raisonner sur le modèle du point de vue. Nous envisageons aussi développé l'aspect évaluation des méthodes et des modèles qui n'est pas traité en profondeur dans ce papier.

3. Weka - A Data Mining Software in Java, <http://www.cs.waikato.ac.nz/ml/weka/>

4. Jena – A Semantic Web Framework for Java, <http://jena.sourceforge.net/>

5. SPARQL - W3C Consortium, <http://www.w3.org/TR/rdf-sparql-ref/>

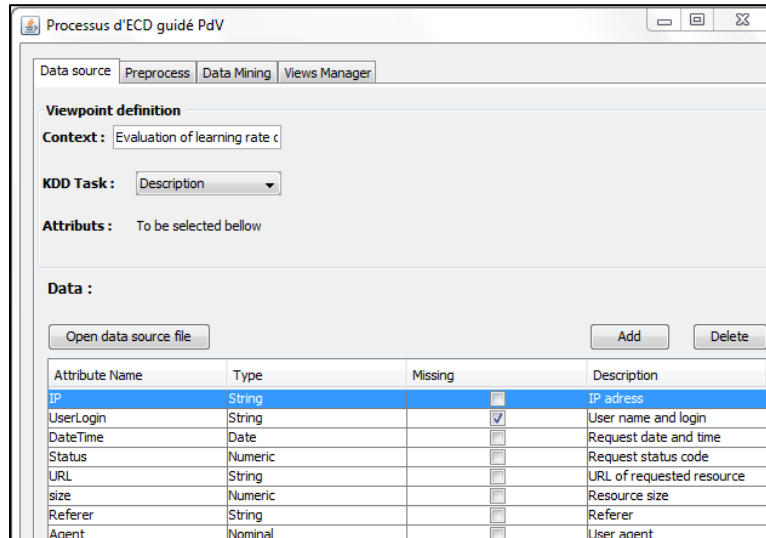


Figure 9. Interface de définition du point de vue d'un enseignant pour analyser les données (fichiers log http) d'un système e-learning

6. Bibliographie

- Aussenac-Gilles N., Laublet P., Reynaud C., *Acquisition et ingénierie des connaissances : tendances actuelles*, Cepaduès Éditions, Toulouse, 1996.
- Bach, T. L., Construction d'un web sémantique multi-points de vue, Thèse de doctorat de l'École des Mines de Paris à Sophia Antipolis, 2006.
- Behja H., Zemmouri E., Marzak A., "Viewpoint-based Annotations for Knowledge Discovery in Databases", *Proc. of IEEE ICMWI*, p. 299-302, Algiers, Algeria, 2010.
- Ben Ahmed W., Safe Next : une approche systémique pour l'extraction de connaissances de données, Thèse de doctorat de l'École centrale Paris, 2005.
- Bernstein A., Provost F., Hill S., "Towards Intelligent Assistance for a Data Mining Process: An Ontology Based Approach for Cost-Sensitive Classification", *IEEE Transactions on Knowledge and Data Engineering* 17(4), p. 503-518, 2005.
- Charlet J., Zacklad M., Kassel G., et Bourigault D., *Ingénierie des connaissances, évolution récentes et nouveaux défis*, Eyrolles, Paris, 2000.
- Diamantini C., Potena D., Storti E., "KDDONTO: an Ontology for Discovery and Composition of KDD Algorithms", *Proc. of the ECML/PKDD09*, p. 13-24, Bled, Slovenia, September 2009.
- Euler T., "Publishing Operational Models of Data Mining Case Studies", *Workshop on Data Mining Case Studies at the 5th IEEE ICDM*, p. 99-106, Houston, Texas, USA, 2005.

- Fayyad U. M., Piatetsky G., Smyth P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data", *Communications of the ACM*, vol. 39, n° 11, p. 27-34, 1996.
- Minsky M., A framework for representing knowledge. Winston, Patrick Henry (Hg.): The psychology of computer vision. New York et al.: McGraw-Hill, p. 211-277, 1975.
- Morik K. and Scholz M., "The MiningMart Approach to Knowledge Discovery in Databases", *Ning Zhong and Jiming Liu (editors), Intelligent Technologies for Information Analysis, Springer*, p. 47-65, 2004.
- Muller A., Caron O., Carré B., Vanwormhoudt G., « Réutilisation d'aspects fonctionnels des vues aux composants », *Revue RSTI-L'objet, LMO'2003, V. 9, n°1-2*, p. 241-255, 2003.
- Nassar M., VUML: une extension UML orientée point de vue, Thèse de doctorat de l'ENSIAS, Université Mohammed V Rabat, 2004.
- Rivière M., Dieng-Kuntz R., "A Viewpoint Model for Cooperative Building of an Ontology", *Actes de ICCS'2002, Springer-Verlag, LNCS 2393*, p. 220-234, 2002.
- Schreiber A. Th., Akkermans J. M., Anjewierden A. A., Hoog R., Shadbolt N., Van de Velde W. et Wielinga B. J., *Knowledge Engineering and Management: The CommonKADS methodology*, MIT Press, Cambridge, Massachusetts, London, 2000.
- Zemmouri E., Behja H., Marzak A., « OntoECD : Une ontologie pour le processus d'ECD », *Workshop WOTIC, Agadir, Maroc, 24-25 Décembre, 2009*.
- Zemmouri E., Behja H., Marzak A., « Architecture web service d'un processus d'ECD multiutilisateurs », *Actes de NGNS'10*, p. 171-177, Marrakech, Maroc, 2010.