



HAL
open science

**Actes du 7e Forum Jeunes Chercheurs du congrès
INFORSID, Lyon, France, 20/05/14 - 23/05/14**

Guillaume Cabanac

► **To cite this version:**

Guillaume Cabanac (Dir.). Actes du 7e Forum Jeunes Chercheurs du congrès INFORSID, Lyon, France, 20/05/14 - 23/05/14. Cabanac, Guillaume. INFORSID (actes électroniques), 2014. hal-03236154

HAL Id: hal-03236154

<https://hal.science/hal-03236154>

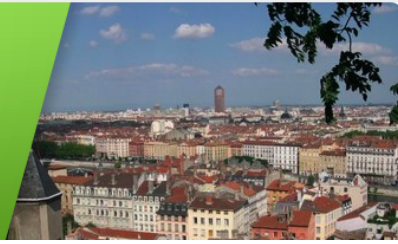
Submitted on 23 Feb 2022

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.

INFORSID 2014

32^e édition - Lyon



Actes du 7^e Forum Jeunes Chercheurs du congrès INFORSID

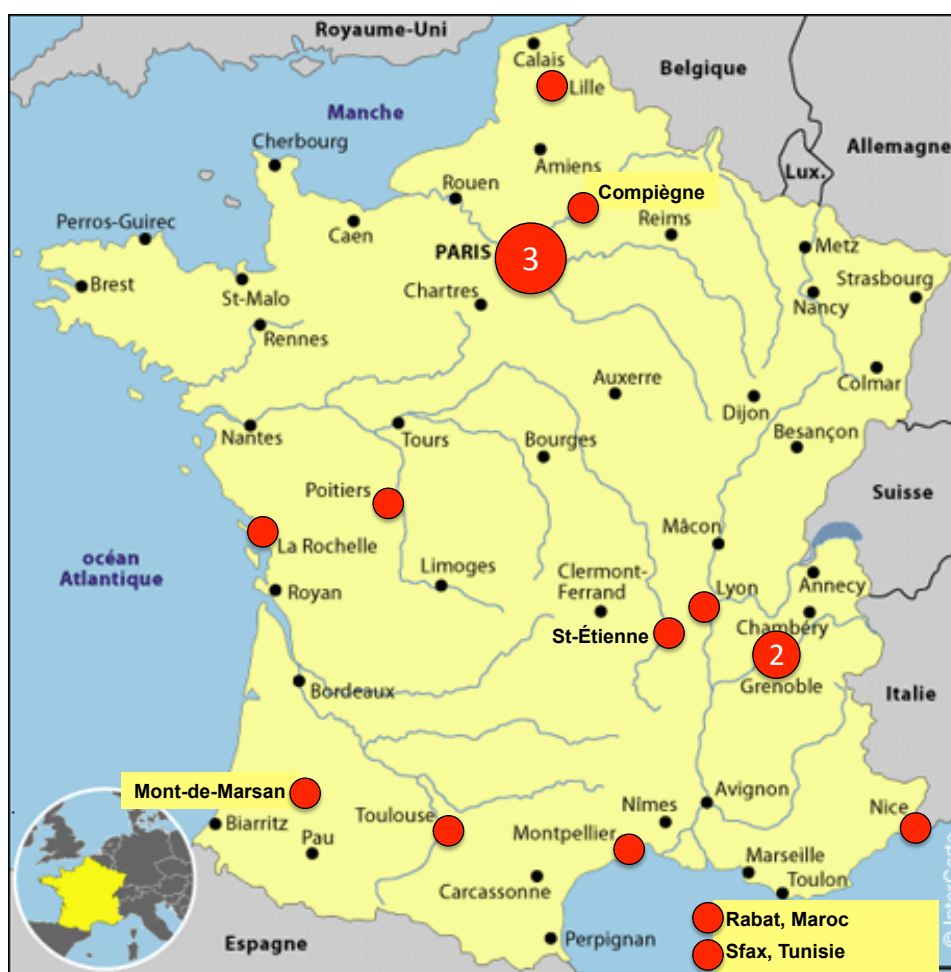
22 mai 2014 à Lyon

Préface

La septième édition du Forum Jeunes Chercheurs est organisée à Lyon le 22 mai 2014, dans le cadre du 32^e congrès INFORSID. L'objectif de cette manifestation scientifique est double :

1. permettre aux jeunes chercheurs en 1^{re} et 2^e année de doctorat de présenter leur problématique de recherche et d'établir des contacts avec des équipes travaillant sur les domaines similaires ou connexes,
2. offrir un aperçu des axes de recherche actuels et ainsi élargir le champ des connaissances des jeunes chercheurs.

Les actes du Forum Jeunes Chercheurs recueillent les 17 articles soumis par des doctorants issus de divers laboratoires de recherche en France, au Maroc et en Tunisie. Ces articles seront présentés lors de la session plénière du congrès consacrée au forum. Durant toute la durée du congrès, l'exposition des posters préparés par les doctorants offrira également des opportunités de rencontres et de discussions avec les chercheurs de la communauté des systèmes d'information.



Répartition spatiale des articles présentés au Forum Jeunes Chercheurs 2014

Le nombre et la qualité des articles ainsi que la variété des thématiques abordées par les doctorants participant au Forum Jeunes Chercheurs sont autant d'éléments qui attestent de la vigueur des recherches dans le domaine des systèmes d'information.

Guillaume CABANAC
Université de Toulouse
Institut de Recherche en Informatique de Toulouse
Organisateur du Forum Jeunes Chercheurs 2014

Table des matières

Articles de doctorants inscrits en 1^{re} année de thèse :

- **Un framework pour la relaxation de requêtes dans les bases de données du Web sémantique**.....1
Géraud Fokou

- **Vers un système d'information collaboratif pour l'analyse des comportements des utilisateurs des réseaux sociaux durant une catastrophe** 5
Imen Bizid

- **Vers une évolution centrée utilisateur des lignes de systèmes d'information**.....9
Eddy Ghabach

- **Vers un modèle de gestion durable du palmier babaçu – Apport de la modélisation de la dynamique de la population**.....13
Nikolay Sirakov

- **Algorithmes de bandits pour les systèmes de recommandation**.....17
Jonathan Louëdec

- **Apport du Web sémantique au travail collaboratif – Application à l'édition d'un document d'information sur les risques majeurs**.....21
Valentin Brun

- **Une approche générique pour la construction collaborative d'éditions critiques numériques**.....25
Vincent Barrellon

- **Vers un système de recommandation à partir de traces sémantiques pour l'aide à la prise de décision**.....29
Ning Wang

- **Construction d'un workflow auto-adaptatif pour le suivi des maladies** 33
Amira Derradji

- **Analyse formelle de concepts pour la gestion de contexte en informatique ubiquitaire**.....37
Ali Jaffal

- **Transformation de modèles de lignes de produits en environnement hétérogène** 41
Farah Maamar

- **Analyse et modélisation de la participation et du E-learning au moyen d'outils collaboratifs**..... 45
Guillaume Blot

Articles de doctorants inscrits en 2^e année de thèse :

- **Conception d'un système expert à base de connaissances floues d'aide à la décision pour la maladie d'Alzheimer.....49**
Firas Zekri

- **Vers un processus d'amélioration continue pour la capitalisation des connaissances à STMicroelectronics..... 53**
Manel Brichni

- **Accords de niveau de service et modèle de réputation pour le contrôle d'usage.....57**
Elena Jaramillo Rojas

- **Application de l'IDM pour la modernisation et la migration d'application existante .. 61**
Khalil Zouaoui

Article collectif de doctorantes inscrites en 1^{re} année et 4^e année de thèse :

- **Vers un enrichissement sémantique des modèles de processus configurables 65**
Loubna El Faquih, Hanae Sbaï

Un framework pour la relaxation de requêtes dans les Bases de Données du Web Sémantique

Géraud Fokou

*LIAS/ISAE-ENSMA et Université de Poitiers
1, Avenue Clement Ader, 86960 Futuroscope Cedex
geraud.fokou@ensma.fr*

MOTS-CLÉS : Relaxation de requêtes, Web Sémantique, Base de Données.

KEYWORDS: Query relaxation, Semantic Web, Database.

ENCADREMENT. Directeurs de Thèse : Allel Hadjali (PR) et Stéphane Jean (MCF)

1. Contexte

RDF est un format de représentation et d'échange de données sous la forme de triplets (*sujet, predicat, objet*) très utilisé sur le Web, comme en témoigne la disponibilité des jeux de données tels que YAGO¹ ou DBPedia². Pour une gestion efficace de ce type de données, des Bases de Données RDF (BDRDF ou triplestore) sont utilisées (par exemple : Oracle Semantic Technology). Elles sont associées au langage SPARQL pour l'interrogation des données.

Une des forces de RDF est sa flexibilité. Il permet en effet de gérer des données qui peuvent être sans schéma explicite ou au contraire décrites par des ontologies définies en termes de classes et de propriétés dans des langages plus ou moins expressifs allant de RDF-Schema jusqu'à OWL Full. De plus, les instances de ce potentiel schéma peuvent avoir des valeurs pour toutes les propriétés (bonne structuration), ou inversement, avoir plus de propriétés sans valeurs (mauvaise structuration) (Duan *et al.*, 2011).

1. YAGO : A large ontology from Wikipedia and WordNet, WWW, 2007

2. DBPedia - A crystallization point for the Web of Data, Journal of Web Semantics, 2009

La flexibilité de RDF, en plus du fait qu'une BDRDF puisse contenir des données intégrées à partir de diverses sources, fait que l'interrogation de ce type de bases de données peut conduire à un résultat de requête qui soit vide. En effet, une requête peut échouer parce que l'on souhaite obtenir la valeur d'une propriété sans savoir qu'elle est peu utilisée dans la source de données interrogée ou parce que l'on interroge une classe qui n'a pas ou a peu d'instances. Une des techniques utilisées pour résoudre ce problème est la relaxation de requêtes qui consiste à élargir une requête qui a échoué pour retourner des résultats alternatifs (best match Vs exact match). Nous nous intéressons dans notre thèse au problème de la relaxation de requêtes dans le cadre des BDRDF.

2. État de l'art

La relaxation de requêtes a été largement étudiée dans le contexte des bases de données relationnelles (Chu *et al.*, 1992). Avec la disponibilité potentielle d'une ontologie au sein de données RDF, de nouvelles possibilités de relaxation sont apparues. Afin de mener une analyse profonde de ces travaux, nous avons considéré les trois critères suivants : les types de relaxation proposés, les mesures de similarité utilisées et la manière dont se déroule le processus de relaxation.

Types de relaxation proposés. La plupart des travaux utilisent deux types de relaxation (Hurtado *et al.*, 2008; Huang *et al.*, 2012) : (1) *les relaxations simples* qui consistent à remplacer une constante par une variable ou à supprimer un triplet dans une requête et (2) *les relaxations conceptuelles* qui utilisent les règles de raisonnement associées au langage d'ontologies utilisé ; elles incluent la relaxation des requêtes par utilisation des hiérarchies de classes et de propriétés. D'autres travaux permettent de remplacer une requête qui échoue par une requête approximative en utilisant une substitution de concepts ou de valeurs dans la requête à réponse vide (Corby *et al.*, 2006; Hogan *et al.*, 2012) ou en modifiant les expressions de chemins (Poulovassilis *et al.*, 2010). Enfin, (Dolog *et al.*, 2009) proposent un processus de relaxation basé sur un profil utilisateur qui permet de remplacer une constante dans une requête par une valeur préférée par l'utilisateur.

Mesures de similarité utilisées. Les mesures de similarité définies dans les différents travaux qui s'intéressent à la relaxation de requêtes sont essentiellement basées soit sur les distances dans les hiérarchies de l'ontologie entre le concept initial et le concept substitué (Corby *et al.*, 2006; Hurtado *et al.*, 2008; Huang *et al.*, 2012) soit sur le nombre d'instances de ces concepts (Huang *et al.*, 2012).

Processus de relaxation. Le processus de relaxation peut être automatique, c'est-à-dire être réalisé sans contrôle de l'utilisateur, en exécutant les requêtes relaxées potentielles par ordre décroissant de similarité avec la requête initiale (Huang *et al.*, 2012). D'autres approches permettent à l'utilisateur d'indiquer la partie de la requête sur laquelle la relaxation doit être effectuée (Corby *et al.*, 2006; Hurtado *et al.*, 2008; Poulovassilis *et al.*, 2010)

3. Problématique

L'état de l'art présenté précédemment nous a permis de faire le constat suivant : (1) peu d'approches permettent de contrôler finement le processus de relaxation, (2) la mesure de similarité utilisée dans ces approches se limite à mesurer la distance entre la requête initiale et la requête relaxée ce qui ne permet pas d'ordonner les instances d'une même requête relaxée, (3) la plupart des techniques de relaxation proposées ne sont pas intégrées au langage SPARQL qui est pourtant le langage standard pour l'interrogation d'une BDRDF et (4) ces approches appliquent un processus de relaxation sans chercher à identifier les causes d'échecs de la requête. En conséquence, nous envisageons de concevoir un framework dans notre thèse qui vise à traiter les problèmes suivants.

Obj1 : l'étude d'un ensemble d'opérateurs de relaxation qui puissent être paramétrés et combinés pour permettre une relaxation progressive et contrôlée de la requête.

Obj2 : la définition d'une mesure de similarité qui permette d'une part de mesurer la proximité entre les résultats alternatifs retournés et la requête initiale et d'autre part de trier ces résultats alternatifs par similarité décroissante pour obtenir les top-k résultats.

Obj3 : l'explication des causes d'échec d'une requête pour une relaxation appropriée.

Obj4 : la prise en compte des préférences utilisateur pour relaxer une requête.

Obj5 : l'optimisation des performances du processus de relaxation.

Obj6 : l'extension du langage SPARQL pour pouvoir utiliser les techniques répondant aux objectifs précédents et l'implantation de ces propositions dans une BDRDF.

4. Actions réalisées

Nous avons défini trois opérateurs pour représenter différents types de relaxation proposés (**Obj1**) : généralisation (*GEN*), substitution de concepts (*SIB*) et relâchement des prédicats (*PRED*). Ces opérateurs permettent une relaxation progressive d'une requête et sont associés à des paramètres permettant de contrôler la relaxation. Ils peuvent être combinés via l'opérateur de conjonction *AND* pour pouvoir faire plusieurs relaxations simultanément. Ces opérateurs ont été formalisés et sont associés à une mesure de similarité qui prend en compte non seulement la similarité entre la requête initiale et la requête relaxée (comme les autres approches) mais aussi le degré de satisfaction des instances retournées par rapport à la requête relaxée (**Obj2**). Ceci permet de discriminer plus précisément les résultats alternatifs retournés.

Au niveau implémentation, les opérateurs définis ont été implémentés comme une extension des langages SPARQL et OntoQL supportés sur la base de données OntoDB. Des tests de performance ont également été réalisés sur un jeu de données réel pour comparer les performances des différents opérateurs et déterminer les conditions d'efficacité de ces opérateurs (répond partiellement aux **Obj5** et **Obj6**).

5. Actions futures

Pour compléter notre framework, nous envisageons de réaliser les actions suivantes.

Cause d'échec des requêtes (Obj3). Le concept de *Minimal Failing Subqueries (MFS)* a été défini par (Godfrey, 1997) pour identifier les requêtes minimales en nombre de prédicats, qui retournent un ensemble vide de résultats. Cette notion est bien adaptée au contexte du Web Sémantique puisqu'une requête est un ensemble de triplets et donc, trouver la cause d'échec d'une requête revient à trouver le(s) triplet(s) qui provoque(nt) cet échec. Par contre, il reste à estimer ces MFS de façon efficace sur des bases de données RDF qui peuvent être volumineuses étant donné que ce problème est connu comme étant NP-complet.

Processus orientés utilisateurs (Obj4). Nous proposerons une modélisation des préférences utilisateurs inspiré des *tags* et/ou des *folksonomies* utilisés dans les systèmes de recommandation. Le stockage et le traitement des préférences utilisateurs dans des bases de données RDF permettront ainsi de guider le relaxation sans aucun feedback des utilisateurs. Une autre proposition serait l'utilisation de l'intelligence collective au service d'un utilisateur. On s'inspirera du *crowdsourcing* pour relaxer les requêtes. L'idée est de mesurer la pertinence d'une réécriture de requête ou d'une approximation de réponse via l'opinion d'un ensemble d'utilisateurs en ligne.

Optimisation du processus de relaxation (Obj5). Le processus de relaxation peut nécessiter l'exécution de plusieurs requêtes qui sont proches de la requête originale. Ceci incite à utiliser des techniques d'optimisation multi-requête qui visent à réduire le temps total d'un ensemble de requêtes en exploitant les parties communes aux différentes requêtes. Ce type d'optimisation est pertinent dans notre contexte car les requêtes partagent des parties communes.

6. Bibliographie

- Chu W. W., Chen Q., Merzbacher M., « CoBase : A Cooperative Database System », *In Demolombe and Imielinski*, Clarendon Press, p. 41–73, 1992.
- Corby O., Dieng-Kuntz R., Faron-Zucker C., Gandon F., « Searching the Semantic Web : Approximate Query Processing Based on Ontologies », *IEEE Intelligent Systems*, vol. 21, n° 1, p. 20–27, janvier, 2006.
- Dolog P., Stuckenschmidt H., Wache H., Diederich J., « Relaxing RDF queries based on user and domain preferences », *JIS*, vol. 33, n° 3, p. 239–260, December, 2009.
- Duan S., Kementsietsidis A., Srinivas K., Udrea O., « Apples and oranges : a comparison of RDF benchmarks and real RDF datasets », *Proceedings of SIGMOD'11*, p. 145–156, 2011.
- Godfrey P., « Minimization in cooperative response to failing database queries », *IJCIS*, 1997.
- Hogan A., Mellotte M., Powell G., Stampouli D., « Towards Fuzzy Query-Relaxation for RDF », *ESWC'12*, p. 687–702, 2012.
- Huang H., Liu C., Zhou X., « Approximating query answering on RDF databases », *Journal of World Wide Web*, vol. 15, n° 1, p. 89–114, January, 2012.
- Hurtado C. A., Poulouvasilis A., Wood P. T., « Query Relaxation in RDF », *JODS*, 2008.
- Poulouvasilis A., Wood P. T., « Combining approximation and relaxation in semantic web path queries », *ISWC'10*, vol. 1, p. 631–646, 2010.

Vers un système d'information collaboratif pour l'analyse des comportements des utilisateurs des réseaux sociaux durant une catastrophe

Imen Bizid

*L3i
Avenue Michel Crépeau
17000 La Rochelle
imen.bizid@univ-lr.fr*

MOTS-CLÉS : Réseaux Sociaux, Système d'Information, Comportement des Utilisateurs, Gestion de catastrophes.

KEYWORDS: Social Networks, Information System, UsersBehavior, Disaster Management.

ENCADREMENT. Patrice Boursier (PR), Sami Faiz (PR), Frédéric Rousseaux (MCF) et Jacques Morcos (MCF)

1. Contexte

Les dernières statistiques des catastrophes qui ont eu lieu durant ces dernières années ont prouvé que le nombre des catastrophes est en train d'augmenter d'une année à une autre tout en provoquant des dégâts irréversibles. Cette augmentation est liée essentiellement au changement des caractéristiques des catastrophes naturelles qui ne sont plus statiques à cause des changements climatiques qui affectent notre planète ce qui rend leur gestion plus compliquée. Par ailleurs, les données déjà acquises lors des anciennes catastrophes ne sont plus suffisantes pour contrôler ces nouvelles caractéristiques exceptionnelles (Bizid *et al.*, 2013). Avec l'émergence des réseaux sociaux tels que Facebook et Twitter, il est désormais possible de collecter en temps réel des informations sur la situation des zones affectées par la catastrophe sans avoir

à se déplacer. Par exemple, durant l'attaque de Boston et l'ouragan Sandy, il a été observé que les informations reliées aux deux catastrophes étaient annoncées dans les réseaux sociaux avant qu'elles soient publiées dans les médias officiels tels que CNN et BBC. Après l'attaque de Boston, le FBI a adopté une nouvelle stratégie de collecte de données consistant à encourager les personnes qui étaient présentes sur le lieu de l'attaque à partager leurs vidéos sur Twitter afin de les aider à identifier les suspects. Par conséquent, ces réseaux doivent être exploités davantage par les intervenants en urgence afin de comprendre la situation des zones affectées en temps réel.

Notre travail a pour objectif de concevoir un système d'information permettant d'analyser le comportement des utilisateurs des réseaux sociaux lors d'une catastrophe. Cette étude permettra aux intervenants en urgence de détecter les informations pertinentes en temps réel, d'avoir une vue globale de tout ce qui circule sur les réseaux et d'agir rapidement dans les régions réclamant une intervention immédiate.

2. État de l'art

Les réseaux sociaux sont utilisés dans toutes les phases de gestion de catastrophes de la prévention à la mitigation. Les principaux axes de recherche autour de cette problématique se résument en trois parties :

1) La détection d'événements consiste à analyser les informations des médias sociaux en temps réel afin de pouvoir détecter l'émergence de publications autour d'un sujet relié à une catastrophe. Plusieurs travaux ont été proposés dans cet axe. Earle *et al.* (2012) ont implémenté un détecteur d'événement qui détecte les augmentations rapides de la fréquence des tweets contenant le mot « tremblement de terre » ou son équivalent dans d'autres langues. Yin *et al.* (2012) ont proposé un module de détection des explosions par la gestion en temps réel des tweets en lançant une alerte quand un incident inattendu est perçu. Pohl *et al.* (2012) ont travaillé sur la détection des sous-événements au cours d'une catastrophe en utilisant les réseaux de neurones.

2) La diffusion des alertes par les réseaux et organismes sismiques est devenue très utilisée dans les médias sociaux. Par exemple, USGS diffuse ses alertes en utilisant deux comptes Twitter : @USGSBigQuakes diffuse d'alertes pour tous les séismes de magnitude 5,5 et plus et @USGSsted diffuse les alertes de magnitudes plus faibles.

3) La connaissance de la situation comprend trois niveaux d'activité à savoir : la perception pour l'extraction des données, la compréhension pour l'analyse des données et la projection pour les visualiser sur une carte géographique ou des graphiques. Plusieurs travaux ont été proposés sur ce sujet, Stefanidis *et al.* (2013) ont proposé un système complet d'extraction des données, d'analyse et de visualisation sur des

cartes géographiques afin de détecter l'émergence des points géospatiaux diffuseurs d'information dans un domaine d'intérêt. Imran *et al.* (2013) ont utilisé les ontologies pour classifier les informations extraites de Twitter dans un cas de catastrophe. Starbird *et al.* (2010) ont proposé un système nommé « Tweak the Tweet » permettant de donner un sens aux tweets à travers la proposition d'une nouvelle syntaxe standard permettant de répondre aux questions qui, quoi et où.

3. Problématique

La plupart des travaux présentés sur le thème des catastrophes ont utilisé des outils d'extraction classiques afin de collecter les données des réseaux sociaux tels que les APIs standards. Mais ces APIs ne sont pas suffisantes pour extraire toutes les données permettant de faire des analyses profondes au sujet d'une catastrophe. De même, les données collectées sont analysées en utilisant des techniques de fouille textuelle qui ne peuvent pas être appliquées en temps réel sur un grand nombre de données et dans des situations critiques comme les catastrophes. Par ailleurs, il n'existe pas de système permettant de détecter et suivre les utilisateurs interagissant sur un sujet donné en temps réel afin de comprendre leurs comportements d'une part et de proposer des analyses adaptées au comportement des utilisateurs et au type de catastrophe d'autre part.

4. Actions réalisées

Notre système actuel est capable d'extraire en temps réel les informations pertinentes partagées sur les réseaux sociaux. Ces informations peuvent être exploitées en temps réel par les intervenants en urgence de manière à avoir une vue globale réelle de tout ce qui se passe à proximité des zones affectées par la catastrophe. Ce système reçoit en entrée la liste des mots clés /ou hashtags identifiant la catastrophe et donne en sortie la liste des publications partagées en temps réel par les utilisateurs interagissant au sujet de la catastrophe. L'originalité de ce système repose sur l'utilisation de systèmes multi-agents et des APIs relatives au réseau social ciblé. Le point fort du système multi-agent proposé repose sur le partage des tâches d'extraction en sous-tâches gérées par plusieurs agents qui sont capables de collaborer entre eux pour atteindre un but en commun. L'infrastructure de notre système se compose de quatre agents :

1) L'agent de recherche (AR) cherche la liste des utilisateurs partageant des informations au sujet de la catastrophe en spécifiant les mots clés et les hashtags reliés à l'évènement.

2) L'agent générateur (AG) reçoit la liste des utilisateurs de AR et affecte un agent d'écoute pour chaque utilisateur détecté.

3) L'agent d'écoute (AE) se met à l'écoute d'un utilisateur qui lui a été affecté par l'AG durant la catastrophe afin d'extraire ses publications en temps réel.

4) L'agent gestionnaire de réseau (AGR) détecte les liens d'amitié entre les utilisateurs partageant des informations au sujet de la catastrophe.

À travers ce système, nous sommes capables de suivre en temps réel les publications des utilisateurs interagissant au sujet de catastrophes et de détecter les relations entre eux.

5. Actions futures

La nature extensible des agents autonomes utilisés dans notre système d'extraction en temps réel favorise son utilisation dans d'autres domaines d'application tels que la détection des événements ou des sous-événements et la recherche des nouveaux hashtags et mots clés reliés à un événement. Dans les actions futures, nous allons enrichir les fonctionnalités des agents afin qu'ils soient capables de faire la différence entre les publications pertinentes de celles qui ne le sont pas. En effet, nous allons appliquer de nouveaux critères pour classer les utilisateurs détectés et leurs publications sans avoir recours aux techniques d'analyse textuelle. Ces critères vont être choisis à travers l'analyse du comportement des utilisateurs interagissant au sujet de la catastrophe.

6. Bibliographie

- Bizid I., Faiz S., Boursier P., Che Mustapha Yusuf J., « Integration of Heterogeneous Spatial Databases for Disaster Management », *7th International Workshop on Semantic and Conceptual Issues in GIS*, Lecture Notes in Computer Science, 2013.
- Earle P., Bowden D., Guy M., « Twitter earthquake detection : earthquake monitoring in a social world », *Annals of Geophysics*, 2012.
- Imran M., Elbassuoni S., Castillo C., Diaz F., Meier P., « Practical extraction of disaster-relevant information from social media », *Proc. of Workshop on Social Media Data for Disaster Management*, WWW '13 Companion, ACM/IW3C2, Republic and Canton of Geneva, Switzerland, p. 1021–1024, 2013.
- Pohl D., Bouchachia A., Hellwagner H., « Automatic sub-event detection in emergency management using social media », *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, ACM, New York, NY, USA, p. 683–686, 2012.
- Starbird K., Stamberger J., « Tweak the tweet : Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting », *7th international ISCRAM Conference*, 2010.
- Stefanidis A., Crooks A., Radzikowski J., « Harvesting ambient geospatial information from social media feeds », *GeoJournal*, vol. 78, n° 2, p. 319–338, décembre, 2013.
- Yin J., Lampert A., Cameron M., Robinson B., Power R., « Using Social Media to Enhance Emergency Situation Awareness », *Intelligent Systems, IEEE*, vol. 27, n° 6, p. 52-59, 2012.

Vers une évolution centrée utilisateur des lignes de systèmes d'information

Eddy Ghabach^{1,2}

¹Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis

²Université Saint Esprit de Kaslik, 446 Jounieh – Liban

³Laboratoire ERIC, Université Lyon 2, 5 avenue Pierre Mendès-France, 69676 Bron Cedex France

ghabach.eddy@etu.unice.fr

MOTS-CLÉS : lignes de produits logiciels, ingénierie de logiciel, feature models, variabilité, agile, reverse engineering.

KEYWORDS : software product lines, software engineering, feature models, variability, agile, reverse engineering.

ENCADREMENT : Mireille Blay Fornarino (PR)¹, Franjeh El Khoury (PR)^{2,3}, Badih Baz (PR)²

1. Contexte

En janvier 2014, le *Gartner group* annonce que les dépenses en matière de «logiciels d'entreprise» connaîtront le taux de croissance annuel le plus important dans le domaine de l'IT, avec des prévisions de 6,8 pour cent¹. En effet, utilisés au quotidien, ces systèmes doivent évoluer en particulier pour améliorer les relations entre l'homme (client, commercial, gestionnaire de logistiques, ...) et la machine, prendre en compte les nouveaux modes de communication et ainsi accroître l'efficacité même des entreprises. La complexité intrinsèque de ces systèmes a conduit à la réalisation de systèmes d'information (SI) « sur étagères » qui doivent aujourd'hui s'intégrer dans des démarches de développement agiles (Navarrete *et al.*, 2007).

Les lignes de produits logiciels visent à supporter la variabilité des applications par l'augmentation de la productivité et de la qualité, et la diminution des coûts et du temps de mise sur marché en s'appuyant sur la réutilisation intensive (Trigaux *et al.*, 2003). Ainsi à l'inverse des méthodes agiles, la construction d'une ligne de produits repose sur la planification d'une réutilisation systématique. Différents travaux ont proposé d'allier agilité et lignes de produits logiciels (Bosch *et al.*, 2011 ; Leitner *et al.*, 2011). Dans la lignée de ces recommandations nous nous intéressons plus spécifiquement à l'évolution des lignes de systèmes d'information en intégrant l'utilisateur dans le processus d'enrichissement.

¹<http://www.gartner.com/newsroom/id/2643919>, 31 mars 2014

2. État de l'art

Une ligne de produits logiciels est un ensemble de systèmes logiciels partageant des propriétés communes, développés sur la base d'un ensemble de composants (Trigaux *et al.*, 2003). En particulier les « feature models » sont une manière de capturer les similitudes et les variations entre les produits de la ligne (Appel *et al.*, 2009). Pour faire face à la complexité croissante des systèmes et à la nécessité d'un développement centré sur les parties prenantes, des écosystèmes logiciels se sont développés. Ils se caractérisent par une décentralisation des développements par des « communautés » de développeurs et l'association des utilisateurs dans le processus même du développement (Bosch *et al.*, 2011). De fait l'évolution des produits issus d'une ligne de SI met en jeu les différents éléments qui composent un SI : base de données, interactions Homme-Machine (IHM), Processus métiers (Authoserre *et al.*, 2012). Les travaux menés dans le cadre des lignes de produits multiples (Bosch, 2010) et de la séparation de préoccupations (Rashid *et al.*, 2011) sont des pistes pour maîtriser les lignes de produits complexes. Notre étude porte actuellement sur leur usage dans le contexte spécifique des SI.

3. Problématique

Les écosystèmes associés aux lignes de produits correspondant à des SI présentent une très grande complexité par leur taille en nombre de composants mis en jeu, par leur variabilité pour adhérer aux besoins spécifiques des entreprises et par la ramification des parties prenantes (Brummermann, H. et al, 2011). De fait, le relevé des exigences des utilisateurs reste une étape difficile, qui peut conduire à des livraisons insatisfaisantes (Bosch *et al.*, 2011). De plus, pour gérer l'évolution, l'approche de développement Top-Down est coûteuse, en temps, voire même en complexité. Nous constatons ainsi sur le terrain l'introduction de nouvelles fonctionnalités directement au niveau des codes. Il s'agit donc d'exploiter l'agilité des développements, la variabilité des composants ajoutés, sans rendre la tâche des développeurs plus complexe. Cela se réalise en appliquant une approche intégrative plus performante en qualité et en rapidité, au niveau de la détermination systématique des variations et de l'enrichissement automatisé de la ligne de produits. La capitalisation des informations dans la ligne est alors une tâche additionnelle et complexe, qui, si elle n'est pas réalisée, peut conduire à l'obsolescence de la ligne qui perd en qualité et surtout en intérêt.

4. Actions réalisées

Dans le contexte de cette problématique, nous avons défini un processus comportant des tâches systèmes et utilisateur. Ces dernières mettent en jeu l'équipe de développement et la maîtrise d'ouvrage représentée par un « Product Owner » (PO). Le processus se décompose en 6 activités, présentées dans la figure 1 :

- L'activité 1 repose sur des techniques classiques de sélection dans des lignes de produits pour choisir les composants nécessaires à un nouveau SI.
- L'activité 2 est opérée si tous les composants requis existent dans la ligne. Les codes sont alors automatiquement générés pour construire l'application et le processus est simplement terminé.
- L'activité 3 consiste à déterminer avec le PO les composants additionnels et ceux à modifier. Cette étape est une de nos contributions. Nous travaillons actuellement à enrichir les approches classiques de sélection de « feature models » par l'introduction de « feature » génériques, en reprenant des principes d'IHMs abstraites (Godet-Bar *et al.*, 2012)
- L'activité 4 génère le modèle partiel de l'application en ajoutant en plus des codes des composants existants, les IHMs vides des nouveaux composants.
- L'activité 5, fondée sur une méthodologie agile de développements, intègre une modélisation dirigée par les IHMs.
- L'activité 6 consiste en une étape de « Reverse Engineering » basée sur des techniques d'intégration des modèles (Mannino, 2004 ; Deltombe *et al.*, 2012), afin de « remonter » les composants ajoutés ou modifiés dans la ligne, avant de générer l'application (cf. étape 2).

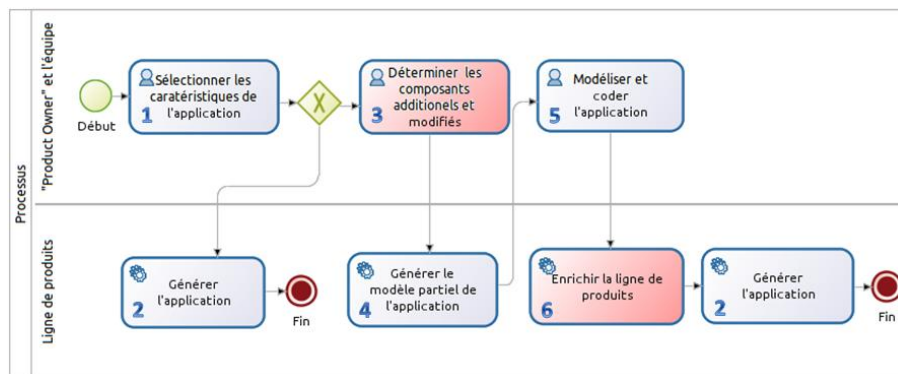


Figure 1 : Processus de développement proposé en BPMN². Les activités colorées correspondent à nos principales contributions.

5. Actions futures

Nous travaillons actuellement sur le moteur de configuration pour prendre en compte en même temps la sélection de features existants et les besoins en nouveaux composants dont des IHMs. Dans un deuxième temps, nous porterons notre attention sur le développement d'une méthode et des outils pour intégrer automatiquement les

² Business Process Model and Notation, <http://www.omg.org/spec/BPMN/2.0/>

nouveaux composants dans la ligne tout en respectant l'intégrité et la variabilité des composants. La structure de la ligne (format des « feature models »), modèles associés aux composants logiciels, devrait nous permettre de définir des opérateurs bien fondés qui garantissent la viabilité de la ligne de produits et plus particulièrement l'intégration automatique des variabilités.

Bibliographie

- Apel, S., Kästner, C., "An Overview of Feature-Oriented Software Development", *Journal of Object Technology*, Vol. 8, July-August 2009, p. 1-36.
- Authoserre, A., Bertrand F., Blay-Fornarino M., Collet P., Dubois H., Ducasse S., Dupuy-Chessa S., Faron-Zucker C., Faucher C., Lafaye J., Lahire P., Le Goer O., Montagnat J., Pinna-Dery M., "Interopérabilité des Systèmes d'Information : approches dirigées par les modèles", *30ème congrès INFORSID Montpellier - France*, 2012, p. 11-30.
- Bosch, J., "Toward Compositional Software Product Lines", *IEEE Software*, May-June 2010, p. 29-34.
- Bosch, J., Bosch-Sijtsema, P.M., "Introducing agile customer-centered development in a legacy software product line", *Software Practice And Experience*, April 2011, p.871-882.
- Brummermann, H., Keunecke M., Schmid K., "Variability issues in the evolution of information system ecosystems", *Vamos'11 Proceedings of the 5th Workshop on Variability Modeling of Software-Intensive Systems*, 2011, ACM, New York, USA, p. 159-164.
- Deltombe, G., Le Goer O., Barbier F., "Bridging KDM and ASTM for Model-Driven Software Modernization", *SEKE. Knowledge Systems Institute Graduate School*, 2012, p. 517-524.
- Godet-Bar, G., Dupuy-Chessa S., Rieu D., "Sonata: Flexible connections between interaction and business spaces", *Journal of Systems and Software*, 2012, p.1105-1118.
- Holl, G., Grünbacher P., Rabiser R., "A systematic review and an expert survey on capabilities supporting multi product lines", *Information and Software Technology*, 2012, p. 828-852.
- Leitner, A., Kreiner, C., "Software Product Lines - An Agile Success Factor?" *Systems, Software and Service Process Improvement*, Vol. 172, Springer, 2011, p. 203-214.
- Mannino, M. V., "Database Design, Application Development, and Administration", *McGraw-Hill/Irwin*, 2004.
- Navarrete, F., Botella, P., Franch, X., "Reconciling Agility and Discipline in COTS Selection Processes", *Commercial-off-the-Shelf (COTS)-Based Software Systems, ICCBSS '07. Sixth International IEEE Conference on*, 2007, p. 103-113.
- Rashid, A., Royer J-C, Rummler A., "Aspect-Oriented, Model-Driven Software Product Lines", *The AMPLE Way*, Cambridge University Press, 2011.
- Trigaux, J.C., Heymans, P., "Software Product Lines: State of the art", *Product Line Engineering of food Traceability software, FUNDP - Equipe LIEL*, Belgique, 2003, p. 9-39.

Vers un modèle de gestion durable du palmier babaçu

Apport de la modélisation de la dynamique de population

Nikolay Sirakov^{a,b}

^aINRA - UMR MISTEA (bât. 29)
2, place Pierre Viala
34060 Montpellier
nikolay.sirakov@supagro.inra.fr

^bUMR ESPACE-DEV
Maison de la télédétection
500, rue J.-F. Breton
34093 Montpellier, France

MOTS-CLÉS : palmier babaçu, dynamique de population, modélisation

KEYWORDS: babassu palm, population dynamics, modeling

ENCADREMENT: Thérèse Libourel^b (PR), Danielle Mitja^b (CR) et Patrice Loisel^a (CR)

1. Contexte

Les travaux relatés dans cet article s'effectuent dans le cadre d'un projet « Open Science » : méthodes et outils pour l'aide à la décision relative à la gestion durable du babaçu dans les pâturages du Brésil, au sein des UMR MISTEA (INRA) et ESPACE-DEV (IRD, UM2, UAG, UR) et concernent des chercheurs thématiques, télédéTECTEURS, mathématiciens, informaticiens et économistes.

Le palmier babaçu (*Attalea speciosa*) est une espèce endémique de l'Amérique du Sud qui est largement représentée au Brésil. En 1982 il couvrait une superficie d'environ 200 000 km² (May *et al.*, 1985). Le babaçu fait partie des ressources brésiliennes dites « extractivistes », c'est-à-dire liées à une activité de cueillette suivie de commercialisation de produits non ligneux de la forêt (Pinton *et al.*, 1992). Dans les années 1980, cette activité concernait un total de 2 millions de personnes parmi les plus démunies du pays (May *et al.*, 1985). Depuis cette période, ces populations majoritairement féminines appelées « Quebradeiras de coco babaçu » (femmes casseuses des fruits du babaçu), sont de plus en plus confrontées à des problèmes d'ordres socio-économique

et géo-politique (Almeida *et al.*, 2005). Ces problèmes résultent de la confrontation entre économie basée sur une agriculture « native » et une économie basée sur une agriculture plus intensive ou sur des secteurs pétro-chimiques. Une prise de conscience émerge dans certains États fédéraux brésiliens et les droits des *Quebradeiras* ont été inscrits dans des lois (Almeida *et al.*, 2005). Elles ont notamment obtenu, dans certaines localités ou États brésiliens, l'accès à toutes les terres sur lesquelles le babaçu est présent y compris les propriétés privées, afin de pouvoir librement récolter les fruits. Le palmier babaçu a également été classé comme espèce protégée : sa coupe est désormais officiellement interdite dans les États de Maranhão, Piauí, Tocantins, Pará, Goiás et Mato Grosso (Almeida *et al.*, 2005).

La recherche que nous menons est donc destinée à aider la prise de décision en étayant les connaissances encore bien imparfaites liées à la dynamique de développement de cette espèce dans les pâturages et au rôle joué par la « cueillette » sur celle-ci.

2. État de l'art

Nous axons cette section autour de l'existant en terme de modèles numériques. La modélisation de populations végétales se base assez souvent sur des modèles matriciels de transition (Liang *et al.*, 2013). Pour de nombreuses plantes, ces modèles s'appuient sur le formalisme de Lefkovitch : classification opérée en fonction de la structure en stades de la population végétale (Logofet *et al.*, 2006). L'avantage de ces modèles est qu'ils peuvent être mis en place relativement facilement par n'importe quel utilisateur et qu'ils sont applicables à la modélisation de toute espèce végétale. C'est pourquoi, durant les trente dernières années, le formalisme de Lefkovitch s'est répandu de manière quasi unanime parmi les spécialistes de différents domaines (Liang *et al.*, 2013) et s'est progressivement démocratisé au sein de la communauté scientifique internationale (Bierzychudek, 1999). Les scientifiques travaillant sur la modélisation de la dynamique de population des palmiers n'ont pas échappé à cet élan et des travaux basés sur les matrices de transition ont vu le jour (Bullock, 1980; Pinard, 1993; Olmsted *et al.*, 1995; Barot *et al.*, 2000; Holm *et al.*, 2008). Ces travaux puisent leur source le plus souvent dans l'ouvrage de référence en matière de modélisation matricielle écrit par Caswell (2001). Le majeur défaut des modèles matriciels basés sur les stades est qu'ils ne peuvent être construits qu'en situation de très légères variabilités annuelles de la dynamique de population. Ces modèles sont fondés sur l'hypothèse que les paramètres mesurés durant la phase de travail de terrain (taux de transition d'un stade à un autre, fécondité, mortalité...) ne vont pas évoluer au cours du temps (Bierzychudek, 1999) : situation stationnaire. Ce postulat est discutable, car il existe une multitude de raisons, qui le rendent inexact : non représentativité de l'échantillon, observations recueillies sur un trop petit laps de temps... (Bierzychudek, 1999). C'est pourquoi, certains auteurs se sont tournés vers des méthodes utilisant conjointement les notions de stade et d'âge des individus (Logofet, 2013). Ce type de modèles fonctionne sur le principe du suivi des cohortes (Cochran *et al.*, 1992). Les modèles matriciels de transition peuvent aussi intégrer de différentes manières la stochasticité environnementale (Kaye *et al.*, 2003).

3. Problématique

Comme souligné précédemment notre objectif est d'aider à la prise de décision et notamment d'apporter quelques réponses à la question qui anime les populations locales : quel modèle de gestion durable du babaçu peut-on envisager ? Cette question couvre plusieurs points qui seront abordés dans la globalité des travaux du projet :

- comment expliciter le fonctionnement durable de cette espèce au sein des milieux récemment anthropisés : pâturages et champs cultivés, suite à l'avancement du front pionnier et de la déforestation des forêts primaires ?

- cela implique de confronter le modèle de dynamique de la population et celui des activités anthropiques ;

- cela implique également de croiser des séries temporelles de données terrain et de données spatiales grâce à l'analyse d'images satellitaires, afin de tester la robustesse des modèles.

4. Actions réalisées

Étant en début de doctorat, les premières étapes du travail ont consisté tout d'abord en l'appropriation du sujet de recherche en réalisant un état de l'art de la modélisation de la dynamique de population des espèces végétales, des palmiers et du babaçu ainsi qu'un état de l'art des différentes techniques et méthodes de traitement d'images satellitaires. Nos réflexions ont ensuite porté sur la proposition d'un premier modèle de l'espèce babaçu. Pour cela, nous avons effectué l'analyse critique du formalisme de Lefkovitch (limites liées à la non prise en compte de la variabilité de la durée des stades pour chaque individu). Nous avons montré que la notion d'âge de l'individu est indispensable et indissociable de celle du stade et déterminé le caractère instationnaire de la dynamique de population du palmier babaçu en fonction des paramètres environnementaux. Parallèlement, le traitement d'une série diachronique d'images satellitaires relatives à la zone étudiée a permis de déterminer l'évolution de la disponibilité théorique du palmier babaçu au sein des pâturages.

5. Actions futures

La suite du travail comporte une phase de complétion de l'état de l'art tant sur les aspects modèles, que sur ceux du traitement des images satellitaires. Le modèle proposé sera affiné en tenant compte des variabilités inter et intra-parcellaire : le choix se portera sur la mise en œuvre d'un modèle hiérarchique de dynamique de population des plantes (formalisme bayésien). Le modèle sera ensuite confronté aux données de terrain (un travail complémentaire au nôtre porte en effet sur un protocole sophistiqué d'observations *in situ*). Le traitement des images satellitaires complètera cette validation du modèle d'une part, en vérifiant l'authenticité de l'information relative quant à la date de mise en place des pâturages (indiquée par les propriétaires) à partir d'une

série diachronique d'images satellitaires Landsat et d'autre part en distinguant, à partir d'images à très haute résolution spatiale (Geoeye), les individus adultes laissés au sein des pâturages (non coupés durant la déforestation) de ceux qui ont atteint l'âge adulte depuis la déforestation.

6. Bibliographie

- Almeida A. W. B., Shiraishi Neto J., Martins C. C., *Guerra ecológica Nos Babaçuais (o processo de devastação dos palmeiras, a elevação do preço de commodities e o aquecimento do mercado de terras na Amazônia)*, MIQCB/Balaios typographia, São Luís-MA. 186 p., 2005.
- Barot S., Gignoux J., Vuattoux R., Legendre S., « Demography of a Savanna Palm Tree in Ivory Coast (Lamto) : Population Persistence and Life-History », *Journal of Tropical Ecology*, vol. 16, n° 5, p. 637–655, 2000.
- Bierzychudek P., « Looking Backwards : Assessing the Projections of a Transition Matrix Model », *Ecological Applications*, vol. 9, n° 4, p. 1278–1287, novembre, 1999.
- Bullock S. H., « Demography of an Undergrowth Palm in Littoral Cameroon », *Biotropica*, vol. 12, n° 4, p. 247–255, décembre, 1980.
- Caswell H., *Matrix population models : construction, analysis, and interpretation*, Sinauer, Sunderland (Massachusetts). 722 p, 2001.
- Cochran M. E., Ellner S., « Simple Methods for Calculating Age-Based Life History Parameters for Stage-Structured Populations », *Ecological Monographs*, vol. 62, n° 3, p. 345–364, 1992.
- Holm J. A., Miller C. J., Cropper Jr W. P., « Population Dynamics of the Dioecious Amazonian Palm *Mauritia flexuosa* : Simulation Analysis of Sustainable Harvesting », *Biotropica*, vol. 40, n° 5, p. 550–558, septembre, 2008.
- Kaye T. N., Pyke D. A., « The effect of stochastic technique on estimates of population viability from transition matrix models », *Ecology*, vol. 84, n° 6, p. 1464–1476, juin, 2003.
- Liang J., Picard N., « Matrix Model of Forest Dynamics : An Overview and Outlook », *Forest Science*, vol. 59, n° 3, p. 359–378, juin, 2013.
- Logofet D. O., « Complexity in matrix population models : Polyvariant ontogeny and reproductive uncertainty », *Ecological Complexity*, vol. 15, p. 43–51, septembre, 2013.
- Logofet D. O., Ulanova N. G., Klochkova I. N., Demidova A. N., « Structure and dynamics of a clonal plant population : Classical model results in a non-classic formulation », *Ecological Modelling*, vol. 192, n° 1–2, p. 95–106, février, 2006.
- May P., Anderson A., Balick M., Frazao J., « Subsistence benefits from the babassu palm (*Orbignya martiana*) », *Economic Botany*, vol. 39, n° 2, p. 113–129, 1985.
- Olmsted I., Alvarez-Buylla E. R., « Sustainable Harvesting of Tropical Trees : Demography and Matrix Models of Two Palm Species in Mexico », *Ecological Applications*, vol. 5, n° 2, p. 484–500, mai, 1995.
- Pinard M., « Impacts of Stem Harvesting on Populations of *Iriartea deltoidea* (Palmae) in an Extractive Reserve in Acre, Brazil », *Biotropica*, vol. 25, n° 1, p. 2–14, mars, 1993.
- Pinton F., Emperaire L., « L'extractivisme en Amazonie brésilienne : un système en crise d'identité », *Cahiers des Sciences Humaines*, vol. 28, n° 4, p. 685–703, 1992.

Algorithmes de bandits pour les systèmes de recommandation

Jonathan Louëdec

Institut de Recherche en Informatique de Toulouse (IRIT)

Institut de Mathématiques de Toulouse (IMT)

UMR5505, UMR5219, CNRS

118 Route de Narbonne

31062 Toulouse, France

Jonathan.Louedec@irit.fr

MOTS-CLÉS : Recherche d'Information, Systèmes de Recommandation, Apprentissage actif, Contexte Utilisateur, Problèmes de Bandits

KEYWORDS: Information Retrieval, Recommender Systems, Active Learning, User Context, Bandits Problem

ENCADREMENT. IRIT: Josiane Mothe (PR) et Max Chevalier (MCF), IMT: Aurélien Garivier (PR)

1. Contexte

Les systèmes de recommandation (SR) ont pour objectif de proposer à l'utilisateur des items (documents, objets, films, musiques, informations, ...) susceptibles de l'intéresser. Deux approches principales sont mises en œuvre : le filtrage basé sur le contenu (FBC) recommande à un utilisateur des items similaires à ceux qu'il a déjà aimés par le passé. Le filtrage collaboratif (FC) recommande les items appréciés par les utilisateurs qui ont auparavant fait des choix similaires à ceux de l'utilisateur. D'autres approches existent : le filtrage démographique se base sur ce que l'on sait de l'utilisateur (âge, données démographiques, sexe, ...); le filtrage communautaire utilise les décisions faites par les contacts de cet utilisateur (cette méthode est notamment utilisée dans les SR sociaux).

2. État de l'art

De nouveaux items et de nouveaux utilisateurs apparaissent au cours du temps, ils doivent être pris en compte par les SR. Le FC et le FBC souffrent de quelques limites bien répertoriées dans les travaux existants. Quels items proposer à un nouvel utilisateur ? Ce problème, connu sous le nom de démarrage à froid utilisateur, est commun aux FC et FBC. Par ailleurs il existe des problèmes spécifiques à chaque type d'approches. Le départ à froid côté item (à qui proposer un nouvel item ?) est une limitation du FC uniquement. Celle-ci peut être traitée en comparant le nouvel item avec les items antérieurs : c'est l'idée du FBC. Au contraire si le contenu d'un item est peu décrit, le FBC ne pourra pas établir de similarité, tandis que le FC pourra le proposer en fonction de l'information obtenue par les actions utilisateurs passées (Montaner, 2003). Ainsi, la combinaison entre FC et FBC peut permettre de corriger les problèmes spécifiques de chaque approche : actuellement, les SR les plus efficaces sont basés sur une approche hybride (Ricci *et al.*, 2011).

Koren (2010) souligne qu'un SR doit par ailleurs être capable de prendre en compte l'évolution des besoins de l'utilisateur : c'est la définition d'un SR dynamique. Deux approches pour ce problème sont l'utilisation d'une fenêtre temporelle, permettant de prendre en compte uniquement les événements les plus récents et la mise en place de méthodes de détection de rupture.

3. Problématique

Dans le contexte de la recommandation nous proposons une approche de traitement séquentielle. Lorsque l'ensemble des interactions des utilisateurs avec le SR jusqu'à l'instant t est connu, se pose la question de quels items proposer à l'instant $t + 1$. Il faut pour cela mettre en œuvre une stratégie pour acquérir de l'information sur les utilisateurs et l'ensemble des items (exploration) tout en assurant que le SR donnera de bons résultats (exploitation). Cette problématique est connue sous le nom de dilemme exploration/exploitation. La stratégie doit pouvoir s'adapter aux flux de données (nouvel utilisateur, nouveau document, évolution des goûts de l'utilisateur). Elle devra également prendre en compte le passage à l'échelle, ce qui pose deux problèmes : comment stocker l'information ? Comment optimiser le temps d'apprentissage ? Nous avons choisi d'étudier l'applicabilité de l'approche statistique « problème de bandits » à l'ensemble des problèmes des SR évoqués plus haut. Dans nos premiers travaux nous nous sommes focalisés sur l'aspect volume de données afin de pouvoir ensuite appliquer les algorithmes de bandits sur des collections réelles.

4. Actions réalisées

Fin 2013, le moteur de recherche russe Yandex a proposé un challenge dont l'objectif était de réordonner les dix premiers résultats de 797 867 requêtes soumises par des utilisateurs différents (<http://www.kaggle.com/c/>

yandex-personalized-web-search-challenge). Nous disposions de plus de 34 millions de requêtes passées et les actions utilisateurs liées : clics, temps passé sur la page cliquée. Chaque terme et chaque site web était remplacé par un identifiant numérique, il n'était donc pas possible de prendre en compte la sémantique des différents éléments. Les actions passées des utilisateurs constituaient les seuls éléments dont nous disposions pour traiter le problème. Dans un premier temps nous avons dû mettre en place une structure permettant de stocker et de traiter rapidement ces données. Nous avons fait le choix d'utiliser MongoDB, un système de gestion de base de données NoSQL orientée documents. Via un script Python, l'insertion en base des données a duré deux heures et, une fois les index créés, la base occupait 117 Go. Nous avons proposé plusieurs stratégies pour répondre au challenge de ré-ordonnement. Celui présenté en Algorithme 1 à permis d'obtenir les meilleures performances, nous avons obtenu la 33-ième place sur 196 participants.

Algorithme 1 : Méthode mise en œuvre pour le ré-ordonnement

```

1 pour chaque requête  $R$  pour laquelle les résultats doivent être réordonnés faire
2   pour chaque document  $D$  de la liste réponse de  $R$  faire
3     Récupérer le nombre de clics de l'utilisateur sur  $D$  avec au moins un
4     terme de  $R$  dans la requête dans les sessions passées (apprentissage)
5   fin
6 Réordonner les documents en fonction de leur score, conserver l'ordre de
7   départ si ex-æquo
8 fin

```

5. Actions futures

Pour aborder l'aspect flux de données, nous voulons mettre en place une méthode prenant en compte le dilemme exploration/exploitation. En effet nous souhaitons pouvoir détecter un changement de goûts ou prendre en compte un nouvel item sans diminuer les performances du SR. Les algorithmes de bandits sont connus pour proposer des solutions à ce dilemme, voir (Bubeck *et al.*, 2012). Cette approche est séquentielle : à partir des t premiers résultats, une décision est prise à l'instant $t + 1$. En fonction de l'item choisi et du résultat obtenu, l'estimateur associé est mis à jour et le résultat stocké. Peu de calculs sont faits à chaque étape, ce qui permet d'obtenir des temps de traitement très courts. Avec un système de stockage et une structure de données adaptés, il est possible d'obtenir des temps de réponses de l'ordre de quelques millisecondes.

Certains d'entre eux, UCB par exemple, utilise des bornes supérieures de confiance à la place des estimateurs. Une conséquence de cette approche optimiste est qu'un nouvel item sera proposé un certain nombre de fois dès son apparition pour estimer sa valeur. Concernant l'évolution des goûts d'un utilisateur et le vieillissement d'un item, il faudrait adapter la méthode de calcul de la borne supérieure pour détecter cette évolution. Pour cela nous pensons utiliser une fenêtre temporelle et les techniques de détection de rupture.

Dans les versions les plus basiques, les algorithmes de bandits traitent chaque item de manière indépendante : un clic utilisateur sur un item modifiera uniquement l'estimateur de cet item. S'il y a de nombreux items à traiter, nous souhaitons inférer de l'information sur les items similaires à celui choisi afin de limiter le temps d'apprentissage. Lorsqu'un nouvel utilisateur apparaît, il s'agit d'acquérir un maximum d'informations en un minimum de temps pour estimer ses goûts. Actuellement les méthodes d'apprentissage actif (Rubens *et al.*, 2011) traitent ce problème en choisissant à chaque instant l'item qui apportera le plus d'informations. L'interaction des méthodes d'apprentissage actif avec celles des problèmes de bandits devrait permettre d'optimiser la partie exploration au profit de l'exploitation.

En recherche d'information, plusieurs chercheurs ont adapté les algorithmes de bandits. Li *et al.* (2010) proposent un algorithme linéaire de bandits, LinUCB, prenant en compte le contexte utilisateur. Dans le cadre du ré-ordonnement d'items, Radlinski *et al.* (2008) proposent le « Ranked Bandits Algorithm ». Les auteurs mettent en place des algorithmes de bandits pour chaque position possible dans l'ordonnement des résultats. Ils prouvent que sur le long terme l'algorithme mis en place est meilleur qu'un algorithme basé uniquement sur la popularité des documents. La méthode d'évaluation reste tout de même très théorique, et une évaluation sur un exemple réel serait intéressant. Pour nos expérimentations, nous utiliserons des jeux de données couramment utilisés dans l'étude des SR. Par exemple le jeu de données MovieLens, contenant dix millions de notes réparties sur 10 681 films et 71 567 utilisateurs, pourrait permettre d'évaluer les aspects flux de données et de passage à l'échelle.

6. Bibliographie

- Bubeck S., Cesa-Bianchi N., « Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems », *Foundations and Trends in Machine Learning*, vol. 5, n° 1, p. 1-122, 2012.
- Koren Y., « Collaborative Filtering with Temporal Dynamics », *Commun. ACM*, vol. 53, n° 4, p. 89-97, avril, 2010.
- Li L., Chu W., Langford J., Schapire R. E., « A Contextual-Bandit Approach to Personalized News Article Recommendation », *Proc. of 19th International World Wide Web Conference*, April, 2010.
- Montaner M., « A Taxonomy of Personalized Agents on the Internet », *Artificial Intelligence Review*, 2003.
- Radlinski F., Kleinberg R., Joachims T., « Learning Diverse Rankings with Multi-Armed Bandits », *Proc. of 25th International Conference on Machine Learning*, 2008.
- Ricci F., Rokach L., Shapira B., « Introduction to Recommender Systems Handbook. », in F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (éd.), *Recommender Systems Handbook*, Springer, p. 1-35, 2011.
- Rubens N., Kaplan D., Sugiyama M., « Active Learning in Recommender Systems », in F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (éd.), *Recommender Systems Handbook*, Springer, p. 735-767, 2011.

Apport du Web sémantique au travail collaboratif

Application à l'édition d'un document d'information sur les risques majeurs

Valentin Brun

*Institut Henri Fayol
École Nationale Supérieure des Mines de Saint-Étienne
158, cours Fauriel CS62362
42023 SAINT-ÉTIENNE cedex 2*

Valentin.Brun@emse.fr

MOTS-CLÉS : Travail collaboratif, Web sémantique, ontologie, métadonnées, risque majeur

KEYWORDS: Collaborative Work, Semantic Web, Ontology, Metadata, Major Risk

ENCADREMENT. Olivier Boissier (PR) et Philippe Beaune (MCF)

1. Contexte

« Collaborer », « travailler ensemble », ce sont des réalités familières à chacun d'entre nous. Or les techniques ont progressé et ont pris un nouveau tournant avec l'arrivée du numérique dans l'entreprise. Bien que l'introduction du numérique ne se fasse pas sans difficultés, de nouveaux outils continuent d'apparaître : le travail collaboratif assisté par ordinateur est progressivement devenu un domaine d'étude à part entière.

Notre recherche s'inscrit dans ce courant et étudie les apports possibles des technologies du Web sémantique au travail collaboratif. Ces technologies, en apportant les concepts d'ontologie, de métadonnées, d'inférence, participent au passage d'un système d'information à un système plus orienté connaissance.

Devant ce vaste champ d'étude, nous avons pour nous guider un cas d'application qui est d'ailleurs à l'origine de ces travaux. Il s'agit du DDRM « Dossier Départemental sur les Risques Majeurs » un document administratif réalisé sous la supervision du préfet et qui compile les informations essentielles concernant les risques naturels et technologiques majeurs connus sur le territoire d'un département. Ce document d'information et de prévention destiné aux citoyens, doit être actualisé tous les cinq ans, étape nécessitant le concours de plusieurs acteurs issus d'horizons et de métiers différents. C'est ici qu'intervient la possibilité d'améliorer ce travail, qui est par essence collaboratif, à l'aide des nouvelles technologies et en particulier celles issues du Web sémantique. Nous imaginons que le système, plongé dans le Web 2.0 et les *Linked Open Data*, favorisera les échanges entre départements.

2. État de l'art

Dans cette section, nous exposons brièvement l'état actuel des connaissances dans différents domaines – choisis comme étant les principales pistes de recherche que nous suivons pour répondre à notre problématique – qui sont l'intégration de données, le Web sémantique et l'annotation sémantique.

L'élaboration d'un DDRM nécessitant de rassembler des données hétérogènes et distribuées, il se pose le problème de l'hétérogénéité des bases de données. Le résumé des différentes approches classiques dans (Hacid *et al.*, 2004), met en lumière les nouveaux enjeux et opportunités ouverts par les travaux du Web sémantique : ce ne sont plus seulement des données mais aussi des rapports techniques, des diaporamas et des fichiers multimédias qui sont manipulés par des services Web.

Afin de gérer l'hétérogénéité sémantique par le biais d'une ontologie, il existe une ébauche d'ontologie sur les risques proposée par (Provitolo *et al.*, 2011). Cette dernière adopte une démarche originale en prenant le parti d'une approche systémique du risque permettant d'embrasser une vision d'ensemble. Ceci autorise l'ontologie d'être modulaire, flexible et évolutive.

Par rapport à une annotation « papier », une annotation numérique, comme le souligne (Wang, 2005), peut recevoir des traitements informatiques : être modifiée, être elle-même annotée, relier d'autres ressources. Les travaux concernant les annotations sémantiques, comme (Sanderson *et al.*, 2013), se regroupent autour de l'*Open Annotation Community Group* du W3C. L'architecture présentée forme une base incontournable pour le développement, comme nous le souhaitons, d'un système d'annotation du Web.

3. Problématique

Concentrons-nous maintenant sur les avancées et les limites des pistes que nous avons citées. Partons des données, puis passons aux documents pour finalement arriver à la dimension « humaine ».

Pour réaliser l'actualisation d'un DDRM, il faut avoir accès à un certain nombre de données qui sont hétérogènes et distribuées. En effet, les données sont récoltées et stockées par des acteurs différents, ce qui implique de rassembler des bases de données différentes, s'appuyant sur des modélisations différentes. Afin d'assouplir cette hétérogénéité, nous souhaitons développer une ontologie du risque majeur facilitant les échanges entre acteurs. Nous pensons également réutiliser des ontologies existantes et communes comme FOAF et SIOC pour modéliser les relations humaines entrant en jeu lors d'un processus collaboratif.

En plus de ces données, de nombreux documents sont manipulés : règlements, rapports, cartes, ... Nous souhaitons améliorer l'utilisation – et en particulier la recherche – des documents en les enrichissant sémantiquement. Cet enrichissement sémantique consiste d'une part à ajouter des métadonnées aux documents – origine, auteurs, mots-clés – et d'autre part, il s'agit de permettre une annotation plus libre des utilisateurs sur les documents. En effet, cette partie permet d'exprimer une opinion qui peut ajouter une plus-value lors de la lecture du document. La mise en rapport de différentes opinions enrichit les points de vue pris en compte dans la réflexion aboutissant à la prise de décision, après une étape de validation sociale comme proposée dans (Cabanac *et al.*, 2010).

Après les données et les documents, nous souhaitons nous pencher sur les parties prenantes : décideurs, experts, utilisateurs, partenaires, ... Pour les faire rentrer dans la boucle, il est nécessaire de capturer leurs discussions et leurs choix à l'intérieur de l'espace numérique collaboratif. Nous pourrions ainsi effectuer des traitements informatiques sur cette activité, dans le but de découvrir des idées réalisées par d'autres personnes dans d'autres projets, dans des endroits et moments différents, mais qui sont pertinentes dans le projet courant car elles partagent avec lui des buts identiques ou similaires. Un exemple pour illustrer cela : deux équipes élaborant un DDRM dans leur département respectif, ont un but identique facilitant l'échange de réflexions entre ces équipes.

Nous concluons en indiquant que les technologies du Web sémantique face à ces défis sont prometteuses. En effet, elles ont l'avantage d'autoriser une démarche évolutive et extensible, d'automatiser certaines tâches et de se coupler relativement bien avec des services existants.

4. Actions réalisées

Nous avons commencé le développement d'une ontologie sur les risques majeurs en partant de l'ontologie modulaire de (Provitolo *et al.*, 2011). Nous avons par la suite commencé à affiner des concepts, comme la taxonomie des aléas, en suivant celle de la maquette nationale du DDRM.

Afin de démontrer la faisabilité de l'architecture proposée, un prototype, JONAS, a été développé. Ce dernier permet d'ajouter, de visualiser et de naviguer sur des métadonnées contenues dans des textes en *Docbook*, ceux-ci formant des fragments du DDRM

en ébauche. Ces métadonnées, exprimées en RDFa, emploient en plus du classique *Dublin Core* pour les métadonnées documentaires, l'ontologie sur le risque majeur précédemment développée et le thésaurus RAMEAU pour les sujets « généralistes » des documents.

Le prototype illustre également les possibilités d'automatisation du renouvellement d'un DDRM. JONAS maintient à jour un DDRM « abstrait » contenant en plus du texte des liens vers des bases de données pour générer des tableaux. Cela permet d'engendrer un document PDF qui sera toujours le plus « à jour ».

5. Actions futures

Le système d'annotation de JONAS a été pensé pour ajouter des métadonnées aux fichiers XML créés par l'application : c'est ce que nous avons précédemment nommé l'enrichissement sémantique objectif. Pour passer à l'enrichissement sémantique subjectif et donc permettre d'exprimer une opinion à l'aide d'une annotation, il est plus intéressant d'opter pour un système qui suit une architecture « *Open Annotation* ». Cela permet d'enregistrer différents points de vue hors du document et de séparer le système d'annotation du reste de l'application. Cette séparation évite de se limiter à des annotations internes ce qui est nécessaire pour annoter toute ressource Web ; et ainsi ne plus être limité à l'annotation d'un texte et envisager l'annotation d'images, cartes, vidéos, diaporamas, ...

Ici apparaissent un certain nombre de défis : comment annoter un fichier dont on n'a que les droits en lecture ? Comment prendre en compte l'évolution des documents dans le système d'annotation afin d'éviter les annotations orphelines ? Comment annoter une sous-partie d'un document ?

6. Bibliographie

- Cabanac G., Chevalier M., Chrisment C., Julien C., « Social validation of collective annotations : Definition and experiment », *Journal of the American Society for Information Science and Technology*, vol. 61, n° 2, p. 271–287, 2010.
- Hacid M.-S., Reynaud C., « L'intégration de sources de données », *Revue Information - Interaction - Intelligence (I3)*, vol. 4, n° 2, p. 1 – 22, juin, 2004. Numéro hors série.
- Provitolo D., Dubos-Paillard E., Müller J.-P. et al., « Vers une ontologie des risques et des catastrophes : le modèle conceptuel », *XVI èmes rencontres de Rochebrune, Du 19 au 23 janvier 2009, Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels*. 1–16, 2011.
- Sanderson R., Ciccacese P., de Sompel H. V., Open Annotation Specification, Community draft, Open Annotation Community Group, février, 2013. <http://www.openannotation.org/spec/core/>.
- Wang S., Annotation persistence over dynamic documents, PhD thesis, Massachusetts Institute of Technology, 2005.

Une approche générique pour la construction collaborative d'éditeurs critiques numériques¹

Vincent Barrellon

*INSA de Lyon, LIRIS
Bâtiment Blaise Pascal
20, Avenue Albert Einstein
69100 Villeurbanne*

vincent.barrellon@insa-lyon.fr

MOTS-CLÉS : Édition électronique, annotation, travail collaboratif, éditeurs bidirectionnels, common ground.

KEYWORDS: Digital publishing, annotation, cooperative work, bidirectional transformations, common ground.

ENCADREMENT: Sylvie Calabretto (PR), Olivier Ferret (PR) et Pierre-Édouard Portier (MCF)

1. Contexte

En maturation depuis plusieurs décennies, les humanités numériques connaissent actuellement un développement important. Définies par leur vocation à constituer une infrastructure numérique de recherche pour la communauté des sciences humaines et sociales, elles rassemblent aujourd'hui un ensemble étendu de projets et d'institutions présentant des missions et des objectifs variés.

Dans cet écosystème, tirant profit de vastes campagnes dédiées à la numérisation des sources patrimoniales, de nombreux projets d'édition électronique ont vu le jour. Ces entreprises illustrent le fait que, au-delà des analogies qu'il est possible de faire entre l'édition traditionnelle (dont les processus conduisent au tirage sur papier) et l'édition numérique (qui aboutit à la publication sur écran), le changement de support induit de

1. Ce travail bénéficie du soutien financier de l'ARC5.

nouveaux types de contenus, de nouveaux modes d'exploitation de ces contenus et des processus de création différents.

Chacune de ces facettes nouvelles intéresse la communauté informatique, dans les domaines de l'ingénierie documentaire, de la gestion de bibliothèques numériques (*digital libraries*) ou de l'étude du travail collaboratif, par exemple.

Notre travail s'inscrit dans ce contexte. Nous nous intéressons spécifiquement à la tâche de production, par une équipe d'éditeurs, de données structurées reflétant un corpus primaire enrichi par un appareil critique, en vue d'un projet éditorial déterminé. Quatre projets d'édition électronique critique d'envergure sont associés à nos réflexions¹. Ces projets mettent en jeu des équipes éditoriales pluridisciplinaires, distribuées et collaboratives. Ils nécessitent une interface de travail dédiée à l'enrichissement critique des données représentant leurs corpus primaires respectifs – corpus à la fois volumineux, composites (intégrant texte, images, audio) et réclamant un éclairage critique multiple.

2. État de l'art

2.1. Modèles d'annotation

Dans le domaine de édition traditionnelle, la notion d'édition critique est indissociable de celle d'annotation. Par analogie, plusieurs définitions de l'annotation *numérique* ont été proposées. Les *Annotation Graphs* (Bird 2001) se présentent à ce titre comme un exemple synthétique et polyvalent. Le modèle requiert que les contenus à annoter soient indexables selon une seule dimension (e.g. le temps pour une ressource audio). Un graphe d'annotation est alors un graphe acyclique orienté, dans lequel les arcs sont étiquetés (les étiquettes constituent une annotation), et dont les nœuds portent éventuellement des indices repérant les bornes des contenus annotés. Le modèle est compatible avec un typage des annotations ; au moyen d'un jeu d'identifiants et de références, il est également possible d'informer une relation entre plusieurs annotations. Agosti et Ferro (2007) formulent une modélisation intéressante des attributs d'une note : ils prévoient en particulier que ces attributs seront évolutifs dans le temps. L'importance donnée à l'évolution des annotations dans le temps revêt un sens différent chez (Gayoso 2012), pour qui des *schémas* d'annotation doivent pouvoir être définis à n'importe quel instant au cours d'un projet d'édition, et pas seulement en amont de l'annotation.

2.2. Common ground

Le *common ground* est un modèle de communication proposé par (Clark 1991), destiné à expliquer comment l'exécution collective d'une tâche est possible malgré une compréhension mutuelle imparfaite et l'impossibilité de s'accorder sur une représentation des objets de l'interaction. La théorie du *common ground* précise que l'interaction

1. Projet de l'Institut Desanti (institutdesanti.ens-lyon.fr) ; projet « Manuscrits de Stendhal » (manuscrits-de-stendhal.org) ; projet ENCCRE (enccre.academie-sciences.fr) ; projet d'édition des dossiers documentaires de *Bouvard et Pécuchet* (dossiers-flaubert.fr).

est rendue possible par le sentiment d'une compréhension mutuelle, suffisante pour les besoins de l'action à effectuer. Bromme (2000) complète en indiquant que l'uniformisation des représentations du problème qui fait l'objet de l'activité collaborative ne doit en aucun cas constituer un objectif : la divergence de points de vue constitue au contraire un moteur dans l'interaction.

3. Problématique

En pratique, les équipes éditoriales associées à notre travail fondent chacun de leurs projets sur une structure de données définie explicitement. Reflet de la politique éditoriale, elle spécifie le vocabulaire et la grammaire d'annotation qui seront instanciés lors de l'édition. La structure de données contraint donc l'expressivité des annotations. Des cas d'utilisation issus des projets associés à ce travail montrent que cette structure de données ne saurait être fixée définitivement. Au contraire, conformément à (Gayoso 2012), elle doit pouvoir être évolutive, c'est-à-dire faire l'objet d'aménagements alors même qu'elle est partiellement instanciée sur le corpus. En outre, les éditeurs individuels prenant part à un projet collaboratif, du fait de la complémentarité de leurs domaines d'expertise, auront éventuellement des besoins en expressivité différents, voire contradictoires (Bromme 2000). Par exemple, tel éditeur cherchera à raffiner une notion générale pour une étude particulière.

La problématique abordée est incarnée par ce paradoxe : la structure de données doit se faire le reflet d'une politique éditoriale cohérente et en même temps répondre aux besoins en expressivité éventuellement antagonistes formulés par les éditeurs individuels.

4. Actions réalisées

Pour résoudre *conceptuellement* le paradoxe précédent, nous avons fait appel à la théorie du *common ground* et à la notion d'éditeur bidirectionnel.

Une structure de données peut être vue comme la mise en forme d'une *représentation* de l'édition en projet. Dans ce cas, (Clark 1991) permet de soutenir l'idée selon laquelle une structure de données pourrait être constituée d'un cœur, comportant des types généraux instanciables par tous, et de structures spécialisées satellites, dynamiquement proposées par les éditeurs en fonction de leurs besoins en expressivité et confrontées les unes aux autres (conformément à Bromme (2000)).

Les éditeurs bidirectionnels (Czarnecki 2009) définissent des mécanismes pour maintenir la cohérence entre des sources d'information structurée partageant des items. Dans notre cas, si un éditeur définit une structure de données alternative par rapport à la structure de cœur, il s'agit de permettre la traduction des données instanciant chaque structure dans l'autre structure ; de garantir que les modifications effectuées dans l'un des domaines soient répercutées dans l'autre domaine ; de gérer les conflits entre des modifications effectuées dans les deux domaines simultanément.

5. Actions futures

La première des tâches que nous nous assignons consistera à approfondir le formalisme que nous esquissons ci-dessous, qui s'appuie sur UnQL+ (Hidaka 2010), modèle de graphe possédant une sémantique bidirectionnelle :

Soit S un schéma représentant une structure de données initiale. Soit I_S une instanciation de ce schéma. Pour des besoins propres, un éditeur va extrapoler un schéma S' à partir de S au moyen de transformations synthétiquement notées g . Dans l'état actuel de nos lectures, nous pressentons qu'un schéma et son instanciation sont équivalents au sens de la bisimulation (Hidaka 2010), ce qui est noté $S \equiv I_S$. Étant donné S, S', g et I_S , nous rechercherons la transformation bidirectionnelle Ff/Bf telle que :

$$\begin{array}{ccc} S & \xrightarrow{g} & S' \\ \equiv & & \equiv \\ I_S & \xleftrightarrow{Ff/Bf} & I_{S'} \end{array}$$

Cette étude formelle débouchera ensuite sur une mise en œuvre expérimentale de la proposition que nous formulons ici, au sein d'une interface d'édition, simplifiée dans un premier temps, soumise aux quatre équipes éditoriales.

Bibliographie

- Agosti, M., Ferro, N. "A formal model of annotations of digital content". ACM Transactions on Information Systems (TOIS), 26(1), 3, 2007.
- Bird, S., Liberman, M. "A formal framework for linguistic annotation". Speech communication, 33(1), 23-60, 2001.
- Bromme, R. "Beyond one's own perspective : The psychology of cognitive interdisciplinarity". Practicing interdisciplinarity, 115-133, 2000.
- Clark, H. H., Brennan, S. E. "Grounding in communication". Perspectives on socially shared cognition, 13, 127-149, 1991.
- Czarnecki, K., Foster, J. N., Hu, Z., Lämmel, R., Schürr, A., Terwilliger, J. F. "Bidirectional transformations : A cross-discipline perspective". In Theory and Practice of Model Transformations, 260-283. Springer Berlin Heidelberg, 2009.
- Gayoso-Cabada, J., Ruiz, C., Pablo-Nuñez, L., Sarasa-Cabezuelo, A., Goicoechea-de-Jorge, M., Sanz-Cabrero, A., Sierra-Rodríguez, J. L. "A flexible model for the collaborative annotation of digitized literary works". In Proceedings of the 2012 Digital Humanities Conference, 190-193, 2012.
- Hidaka, S., Hu, Z., Inaba, K., Kato, H., Matsuda, K., Nakano, K. "Bidirectionalizing graph transformations". In ACM Sigplan Notices, Vol. 45, No. 9, 205-216. ACM, 2010.

Vers un système de recommandation à partir de traces sémantiques pour l'aide à la prise de décision

Ning Wang

*UTC -Université de Technologie de Compiègne
HEUDIASYC -Centre de Recherches Royallieu
Rue Roger Couffon
60203 Compiègne CEDEX
ning.wang@utc.fr*

MOTS-CLÉS : Trace d'Interaction, Système de Recommandation, Modélisation Sémantique.

KEYWORDS : Trace of Interaction, Recommender System, Semantic Modeling.

ENCADREMENT : Marie-Hélène Abel (MCF-HDR), Jean-Paul Barthès (PR), Elsa Negre (MCF).

1. Contexte

La collaboration peut être définie comme un processus partagé dans lequel deux personnes ou plus avec des compétences complémentaires interagissent pour créer une compréhension commune (Schrage 1990). L'utilisation d'un environnement informatique peut faciliter la collaboration. La plate-forme web E-MEMORAe2.0 vise cet objectif (Abel 2009). Utiliser un tel environnement permet également d'exploiter les traces d'interactions. Dans ce contexte, une trace est définie par (Zarka 2011) comme un enregistrement des actions effectuées par un utilisateur sur un système. Aujourd'hui de nombreux travaux s'intéressent à l'exploitation de telles traces à différentes fins comme l'aide à la décision, la recommandation, etc. (Li 2013) s'est notamment intéressé aux traces de collaboration.

2. État de l'art

De nombreux travaux s'intéressent à l'exploitation des traces d'interaction (Djouad 2008). Nous retenons particulièrement ceux de (Li 2013) qui permettent de distinguer 4 types de traces : Privées, Collaboratives, Collectives et Personnelles. Une trace privée est émise et reçue par un utilisateur identique. Une trace collaborative a un seul émetteur et plusieurs récepteurs. Une trace collective a

plusieurs émetteurs et plusieurs récepteurs et une trace personnelle a un seul émetteur sans limite de récepteur. Ce modèle autorise une analyse plus fine des interactions des utilisateurs. Il met notamment l'accent sur les échanges entre membres utilisateurs de l'environnement informatique. Une des principales utilisations des traces concerne les systèmes de recommandation.

L'intérêt pour les systèmes de recommandation est conforté par la quantité d'informations à traiter et la nécessité d'avoir une aide pour alléger ce travail (Adomavicius 2005). Les systèmes de recommandation sont généralement classés en 3 catégories : les approches basées sur le contenu, celles sur le filtrage collaboratif et les approches hybrides (Balabanović 1997).

3. Problématique

Dans le cadre de nos travaux, nous nous intéressons à l'apport de modèles sémantiques pour l'exploitation des traces d'interaction à des fins de recommandation. Nous avons choisi de reprendre la typologie de modèles de traces définie par (Li 2013) afin d'établir des recommandations adhoc. Nous distinguons ainsi des recommandations de groupe (amélioration de la collaboration, identification des risques, opportunités, etc. d'un ensemble d'utilisateurs réunis en groupe), individuelles (comment améliorer l'efficacité, l'organisation d'un utilisateur dans ses différentes tâches), privées (comment améliorer l'organisation privée d'un utilisateur), collectives (comment améliorer la communication au sein d'une organisation, etc.).

Le système de recommandation visé tiendra également compte d'un modèle sémantique (Qui travaille avec qui sur quoi ? Le sujet A est plus proche du sujet B, etc.) et des traces d'interactions enregistrées (qui a partagé un document sur le sujet S avec qui ? Qui échange souvent avec l'expert Paul ? etc.).

4. Actions réalisées

Nous avons défini les principaux concepts liés à notre travail de la manière suivante :

- Action : interaction ou acte réalisé par l'utilisateur dans l'environnement collaboratif, par exemple l'envoi d'un document ;
- Trace classée : une suite d'actions réalisées sur l'environnement informatique par un utilisateur, qui est classée selon les 4 modèles proposés par (Li 2013) ;
- Ensemble de traces : ensembles de traces classées.

Nous avons alors fait le choix d'étudier l'apport de l'approche TF-IDF en faisant l'analogie entre mot & action, document & trace classée et corpus & ensemble des traces. Le TF-IDF (de l'anglais Term Frequency-Inverse Document Frequency) (Spärck Jones 1972) est une méthode souvent utilisée en recherche d'information et

en particulier dans la fouille de textes. Typiquement l'approche TF-IDF étudie les relations entre les mots, les documents et le corpus. Si un mot apparaît plus dans un document, et en même temps, il apparaît moins dans d'autres documents de ce corpus, il peut mieux représenter le document. Nous définissons l'équation de TF-IDF adaptée à notre cas comme suit :

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$idf_i = \log \frac{|P|}{|\{p: t_i \in p\}|}$$

Où :

- $n_{i,j}$: le nombre d'actions concernant le sujet i réalisées par l'utilisateur j ;
- $\sum_k n_{k,j}$: le nombre d'actions concernant tous les sujets réalisées par l'utilisateur j , où $n_{k,j}$ est le nombre d'actions concernant le sujet k réalisées par l'utilisateur j , qui comprend aussi $n_{i,j}$;
- $|P|$: le nombre d'utilisateurs dans un groupe ;
- $|\{p: t_i \in p\}|$: le nombre d'utilisateurs dans ce groupe qui ont réalisé au moins une action concernant le sujet i .

La compétence est regardée comme la valeur : $C_{i,j} = tf_{i,j} \times idf_i$.

La figure 1. illustre un exemple d'interaction de deux groupes. Dans chaque groupe, une ligne représente la trace collaborative d'un membre du groupe concernant un sujet précis. Les différents sujets sont distingués par couleur.

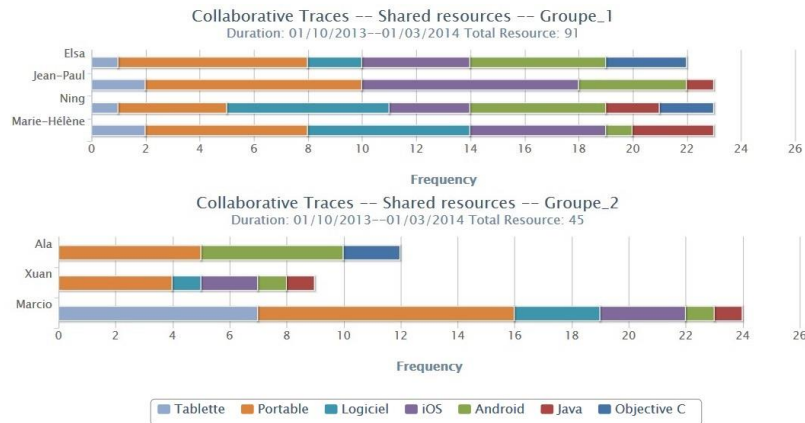


Figure 1. Un exemple de trace collaborative de deux groupes

Le système doit évaluer la compétence de Ning sur le sujet <<Objective C>>. Ning a réalisé 23 actions dont 2 actions concernant << Objective C >>. Dans le groupe 1, le nombre d'utilisateurs est 4 dont 2 ont réalisé au moins une action concernant << Objective C >>. Selon notre méthode d'évaluation, nous obtenons :

$$C_{Objective\ C, Ning} = \frac{2}{23} \times \log \frac{4}{2} = 0.0262$$

De même, nous évaluons la compétence de Ala sur « Objective C » :

$$C_{Objective\ C, Ala} = \frac{2}{12} \times \log \frac{3}{1} = 0.0795$$

Bien qu'ils soient dans deux groupes, nous pouvons déterminer que Ala est plus compétent que Ning sur « Objective C » car $C_{Objective\ C, Ala} > C_{Objective\ C, Ning}$.

5. Actions futures

Dans un environnement collaboratif la date à laquelle une action est réalisée est aussi enregistrée et devrait permettre de prendre en compte le temps : une action ancienne a moins de poids qu'une action récente. Une façon d'exploiter le modèle sémantique serait de définir des poids en fonction de distances sémantiques. Par exemple si nous avons besoin de trouver un expert sur le sujet « Ordinateur », bien qu'aucun utilisateur n'ait laissé de trace sur ce sujet, certains peuvent avoir abordé très fortement d'autres sujets comme « Tablette » sémantiquement proches d'« ordinateur ». Il est donc nécessaire de tenir compte de cette information. Nous prévoyons de pondérer positivement sa compétence sur « Ordinateur ».

Bibliographie

[Abel 2009] M.H. Abel, A. Leblancet al. *Knowledge Sharing via the E-MEMORAE2.0 Platform*. In Proceedings of 6th International Conference on Intellectual Capital, Knowledge Management & Organisational Learning, pages 10-19, 2009.

[Adomavicius 2005] G.Adomavicius, & A. Tuzhilin (2005). *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions*. *Knowledge and Data Engineering*, IEEE Transactions on, 17(6), 734-749.

[Balabanović 1997] M. Balabanović and Y. Shoham, *Fab: Content-Based, Collaborative Recommendation*, Comm. ACM, vol. 40, no. 3, pp. 66-72, 1997.

[Djouad 2008] Djouad, Tarek. "Analyser l'activité d'apprentissage collaboratif: Une approche par transformations spécialisées de traces d'interactions." 2ième rencontre des jeunes chercheurs RJC-EIAH08 (2008): 93-98.

[Li 2013] Qiang LI. *Modeling and Exploitation of the Traces of Interactions in the Collaborative Working Environment* [D]. Thèse Doctorat: UTC -Université de Technologie de Compiègne, 2013.

[Schrage 1990] Michael Schrage. *Shared minds: the new technologies of collaboration*. Random House Inc., New York, NY, USA, 1990.

[Spärck Jones 1972] Jones, K. S. (1972). *A statistical interpretation of term specificity and its application in retrieval*. *Journal of documentation*, 28(1), 11-21.

[Zarka 2011] R. Zarka, A. Cordier, E. Egyed-Zsigmond and A. Mille. *Trace replay with change propagation impact in client/server applications*. In Conference AFIA, 2011.

Construction d'un workflow auto-adaptatif pour le suivi des maladies

Amira Derradji

Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France

CNRS, LIG, F-38000 Grenoble, France

amira.derradji@imag.fr

ARCAN SYSTEMS

12B rue Professeur Deperet

69160 TASSIN LA DEMI LUNE

amira.derradji@arcan.fr

MOTS-CLÉS : processus métier adaptatif, workflow, workflow adaptatif, e-santé.

KEYWORDS: adaptive business process, workflow, adaptive workflow, e-health.

ENCADREMENT: Agnès Front (MCF-HDR), Christine Verdier (PR) et Serge Chvetzoff (collaborateur industriel)

1. Contexte

Notre recherche s'inscrit dans le cadre de la prise en charge médicale à domicile de personnes atteintes de maladies chroniques. Ces dernières qui se caractérisent par une longue phase pathologique et un profil évolutif, peuvent engendrer une perte d'autonomie chez les patients atteints provoquant ainsi une diminution de leurs capacités, d'où la nécessité d'apporter des solutions technologiques permettant d'améliorer leur autonomie. Depuis plusieurs années, diverses pistes de prise en charge ont vu le jour pour mieux accompagner le patient au quotidien. L'Hospitalisation À Domicile (HAD) ou encore le Service de Soins Infirmiers À Domicile (SSIAD) permettent au patient d'être soigné dans son environnement familial et aux professionnels de santé d'effectuer un suivi régulier à distance via des dispositifs installés à domicile.

Plus précisément, le vécu quotidien avec la maladie aide le patient à mieux connaître sa maladie et à en tirer une expertise. Il devient donc une partie intégrante de ses propres soins. Nous pouvons le considérer aujourd'hui comme « *un patient expert* ».

Nos travaux de recherche visent donc à concevoir des protocoles auto-adaptatifs de soins intégrant l'expertise du patient.

2. État de l'art

Beaucoup d'outils mettant en place la prise en charge des patients à domicile sont apparus au cours des dernières décennies. Certains sont centrés sur *la gestion des activités de soins réalisées par les professionnels de santé* (comme la solution *Athome* de notre collaborateur industriel *Arcan Systems*) et d'autres orientés *patients* sont basés sur les services de la e-santé (télé-consultation : service GLUCONET (Perréal, 2003), télé-assistance : projet TelePat (Lacombe et al., 2006), télé-surveillance : projet e-vital (Sakka et al., 2004) et télé-expertise : réseau TELURGE (Desrués, 2008)).

Nous focalisons notre recherche sur (i) la technologie de workflow (WFMC, 1999) pour construire et automatiser le processus de soins, (ii) et les différentes approches d'adaptation des processus (Andonoff et al., 2012) pour adapter le processus de soins suivant les besoins imprévus de chaque patient.

Les workflows sont utilisés dans le domaine de la santé dans le but de répondre à la question d'*auto-adaptabilité* des processus de soins, qui peuvent changer d'un patient à un autre même pour une pathologie identique. Face à cela, le système de gestion de workflow doit être en mesure de changer le processus de soins initial suivant les événements y compris ceux inattendus rencontrés pendant son exécution.

3. Problématique

Selon l'Organisation Nationale de la Santé (OMS), l'éducation thérapeutique du patient vise à aider les patients à acquérir ou maintenir les compétences dont ils ont besoin pour gérer au mieux leur vie avec une maladie chronique (OMS, 1996). Cependant, force est de constater que le patient reste encore dans le rôle d'utilisateur du dispositif qui lui est fourni pour être assisté à distance et ne s'est pas impliqué d'une manière directe dans les soins.

Par ailleurs, les établissements de soins et les organisations sanitaires ont besoin d'outils de suivi de maladies chroniques, qui assurent un accompagnement efficace en dehors du cadre clinique sans (quasi)-intervention médicale et sans désengagement de l'équipe de soins de l'assistance du patient.

L'idée générale de notre travail consiste à mettre en évidence l'expertise et le savoir du patient chronique afin d'améliorer son suivi médical à l'extérieur du milieu hospitalier, en lui offrant la possibilité de signaler tout événement spécifique. L'équipe médicale pourra ensuite interpréter la situation, et mettre en œuvre une action ciblée.

Cet échange d'informations entre le patient et l'équipe médicale se fera via un workflow automatisé qui adaptera les réponses médicales d'une manière quasi-automatique.

Cet outil présentera également un intérêt réel pour l'amélioration de la connaissance médicale car toutes les informations renseignées par le patient seront recueillies et restituées pour (i) répondre à des nouvelles situations pathologiques et (ii) aider le professionnel de santé à mieux s'informer.

4. Actions réalisées

Nous avons dirigé nos recherches autour de deux axes principaux. L'un concerne les processus traditionnels (statiques), définis dans un environnement plus au moins stable. L'autre concerne les processus adaptatifs (dynamiques), qui contrairement au premier type ne peuvent pas être spécifiés à l'avance à cause des changements prévus ou imprévus qu'ils peuvent subir dans leurs environnements de déploiement, aussi bien dans la phase de conception que d'exécution.

Dans le premier axe, nous nous sommes intéressés à la modélisation d'un processus statique correspondant au processus de soins « surveiller la dialyse » (La dialyse est une technique d'épuration permettant d'éliminer les déchets toxiques accumulés dans le sang et dans l'eau en excès. Elle est recommandée pour traiter l'Insuffisance Rénale Chronique). Celui-ci a été modélisé avec différentes approches de modélisation à savoir : UML (Morley et al., 2006), BPMN (OMG, 2010), le diagramme EPC (Davis, 2001), le logigramme, les schémas de processus (Brandenburg et al., 2008) ainsi que les diagrammes de description (Noyé, 1997) dans le but de décrire graphiquement les tâches et les interactions à effectuer par les différents acteurs. Deux approches nous ont permis de représenter les divers événements correspondant aux besoins du patient : BPMN et les diagrammes EPC. À cela s'ajoute la possibilité de modéliser les sous-processus de soins sous BPMN. Donc, le langage BPMN semble le plus approprié pour la représentation des événements prévus liés au processus statique.

Quant au deuxième axe, un approfondissement a été mené sur le processus dynamique où nous avons abordé les trois principaux types d'adaptation (Andonoff et al., 2012) : l'adaptation par flexibilité, qui a pour objectif de retarder la modélisation ou l'exécution du processus, l'adaptation par déviation, qui traite les imprévus pendant la phase d'exécution et l'adaptation par évolution, qui analyse l'exécution afin d'apporter des améliorations à posteriori. Nous avons également exploré les six techniques d'adaptation des processus métier. Il s'agit de l'adaptation par évolution, l'adaptation par cas, l'adaptation par conception, l'adaptation par contrainte, l'adaptation par intention et contexte et l'adaptation par intégration de protocoles.

5. Actions futures

Nous choisirons une des techniques d'adaptation des processus afin de l'étendre aux problèmes ciblés : (i) intégration de l'expertise du patient dans le workflow, (ii)

adaptation du workflow de façon semi-automatique et (iii) reconfiguration à la volée du processus.

Bibliographie

- Andonoff E., Hanachi C., Nurcan S., “*Adaptation des processus d’entreprise*”, L’adaptation dans tous ses états (Philippe Lopisteguy, Dominique Rieu, Philippe Roose (Eds.)), Editions Cépaduès, n° 3, octobre 2012, p. 78-122.
- Brandenburg H., Wojtyna J.P., L’approche processus : mode d’emploi n° 2, Paris, Editions d’organisation, 2008.
- Davis R., Business Process Modelling with Aris : A Practical Guide, London, Editions Springer-Verlag, 2001.
- Desrues L., La télémédecine, un outil au service de l’amélioration du système de santé, L’exemple du réseau TELURGE dans le Nord - Pas-de-Calais, Mémoire de fin d’année, École des hautes études en santé publique, 2008.
- Lacombe A., Boudy J., Baldinger J.L., Delavault F., Muller M., Farin I., An-dreao R. V., Torres-Muller S., Serra A., Gaiti D., Rocaries F., Dietrich CH., Steenkeste F., Schaff M., Baer M., Ozguler A., and Vaysse S., “*Telemedecine for elderly patient at home : the telepat project*”, Smart homes and beyond : ICOST2006, vol. 19, n° 6, 2006, p. 74-81.
- Morley C., Hugues J., Leblanc B., Hernandez J.A., Joly R., Gardian G., UML 2 pour l’analyse d’un système d’information : Le cahier des charges du maître d’ouvrage n° 3, Paris, Editions Dunod, 2006.
- Noyé D., and Mouvement français pour la qualité, L’amélioration participative des processus, Editions INSEP, 1997.
- OMG., Business Process Model and Notation (BPMN 2.0), rapport technique n° 2, Object Management Group, juin 2010.
- OMS., Therapeutic Patient Education – Continuing Education Programmes for Health Care Providers in the field of Chronic Disease, Rapport de l’OMS-Europe, 1996 (traduit en français en 1998).
- Perréal F., and France Télécom Recherche & Développement., “Gluconet : Télémédecine et suivi thérapeutique du patient diabétique”, *12èmes Journées des Technologies de Communication Hospitalière*, Montpellier, 19-20 juin 2003.
- Sakka E., Prentza A., Lamprinos I. E., Leondaridis L., and Koutsouris D., “*Integration of monitoring devices in the e-vital service*”, IEEE Eng Med Biol Soc, vol. 4, n° 26, 2004, p. 097-100.
- WFMC., WfMC workflow Management Coalition Terminology & Glossary, Document Number WFMC-TC-1011, février 1999.

Analyse formelle de concepts pour la gestion de contexte en informatique ubiquitaire

Ali Jaffal

Centre de Recherche en Informatique, Université Paris 1 Panthéon-Sorbonne
90, rue de Tolbiac,
75013 Paris

ali.jaffal@malix.univ-paris1.fr

MOTS-CLÉS : INFORMATIQUE UBIQUITAIRE, SYSTÈMES SENSIBLES AU CONTEXTE, GESTION DE CONTEXTE, RECOMMANDATION.

KEYWORDS : UBIQUITOUS COMPUTING, CONTEXT-AWARE SYSTEMS, CONTEXT MANAGEMENT, RECOMMENDATION.

ENCADREMENT : BÉNÉDICTE LE GRAND (PR) ET MANUELE KIRSCH PINHEIRO (MCF)

1. Contexte

Mon sujet de recherche vise à gérer et analyser les informations liées au contexte d'un utilisateur, disponibles grâce à l'informatique ubiquitaire. D'énormes volumes de données très diverses sont en effet collectés par des capteurs, fournissant par exemple des indicateurs sur la localisation géographique d'un individu, les conditions météorologiques, etc. Les collectes de données de ce type se multiplient, par exemple dans le cadre des villes intelligentes. Ce contexte « environnemental » peut être enrichi d'autres informations relatives aux utilisateurs, par exemple, leurs activités. Ceux-ci peuvent aussi être eux-mêmes fournisseurs d'indicateurs lorsqu'ils renseignent leur profil. L'objectif de mes travaux est de pouvoir prédire ses actions futures en fonction des évolutions de ce contexte ou de lui faire des recommandations.

2. État de l'art

Apparue au début des années quatre-vingt-dix à travers notamment les travaux de Schilit et Theimer (1994) la sensibilité au contexte a été définie comme la capacité d'une application à découvrir et à réagir aux modifications de l'environnement où se trouve l'utilisateur (Dey, 2001). Au centre de ces systèmes se trouvent donc les mécanismes de gestion de contexte (Dey, 2001; Conan *et al.*, 2007; Brézillon, 2003), responsables de l'acquisition et de la représentation de l'information de contexte à partir de l'environnement. C'est à partir de cette information que les processus d'adaptation caractéristiques des systèmes sensibles au contexte peuvent prendre place.

L'information de contexte est complexe, du fait de son caractère dynamique et incomplet, voire incertain (Chalmers *et al.*, 2004; Vanrompay *et al.*, 2011). Par ailleurs, l'acquisition des informations dans un tel environnement est sujette à des imprécisions, du fait notamment de limitations des équipements (capteur défaillant) ou des techniques de collecte utilisées.

La notion de recommandation associée à la notion de contexte n'est pas très détaillée dans la littérature et est souvent ambiguë. Pignotti *et al.* (2004) parlent de recommandation adaptative et présentent le contexte comme un élément majeur pour ce type d'adaptation. Par contexte, les auteurs entendent la date, l'emplacement et l'historique des services utilisés par l'utilisateur. Wen et He (2006) introduisent un système de recommandation personnalisée de services, qui comporte un ensemble d'outils permettant de trouver des informations ou des services intéressants pour l'utilisateur. Chaari *et al.* (2009) relie la recommandation à la personnalisation et à l'adaptation. Le concept de base de ce travail est ainsi de générer automatiquement du code pour l'interface utilisateur pour fournir des services personnalisés en adaptant les offres, en recommandant certains produits, ou en offrant un accès facile à quelques services relatifs à la localisation de l'utilisateur.

Un autre exemple est donné par Najar *et al.* (2012), qui proposent un mécanisme de prédiction cherchant à anticiper les services à invoquer pour un utilisateur en fonction de son contexte. Sigg *et al.* (2010) proposent quant à eux de compléter la description d'une situation avec les éléments de contexte manquants à partir de l'analyse des contextes précédemment observés. Ces travaux utilisent majoritairement des méthodes statistiques et des mesures de similarité pour effectuer une classification des données.

3. Problématique

Bien que les approches mentionnées fournissent des résultats intéressants, elles présentent certaines limites. Les méthodes de classification sous-jacentes ne permettent notamment de générer des classes recouvrantes. En d'autres termes, les éléments de contexte ne peuvent pas appartenir simultanément à plusieurs classes, alors que, dans la pratique, un même élément de contexte peut caractériser différentes situations. Par ailleurs, pour être efficaces, ces méthodes requièrent un volume important de données, qui ne sont pas toujours disponibles. Enfin, la justification des classes identifiées n'est pas forcément compréhensible pour les utilisateurs finaux, ce qui peut conduire à des erreurs d'interprétation et donc de mauvaises recommandations.

Les questions auxquelles nous souhaitons répondre sont notamment : « *Comment identifier les informations de contexte pertinentes ?* », « *dans quel contexte les actions sont-elles réalisées ?* » et « *Quel est l'impact de ces informations sur le choix (et donc la recommandation) d'une action par (ou pour) un utilisateur ?* ».

4. Actions réalisées

Notre objectif est d'exploiter au mieux les données disponibles afin de comprendre le comportement de l'utilisateur et éventuellement faire des recommandations. Nous proposons une méthodologie qui repose sur l'Analyse Formelle de Concepts (AFC) (Priss, 2006; Wille, 2005) et les règles d'association (Agrawal et al., 1993; Agrawal et Srikant, 1994) pour la gestion de contexte à des fins de recommandation. Cette méthode permet de découvrir des relations non triviales entre les éléments de contexte et des actions apparemment décorrelées et d'évaluer l'impact des éléments de contexte sur les actions de l'utilisateur. L'AFC est une méthode mathématique pour l'analyse de données qui permet le regroupement conceptuel par des treillis de concepts, également appelés treillis de Galois. Grâce à l'AFC, nous pouvons organiser les informations contextuelles observées sur des classes pouvant se recouvrir, quelle que soit la granularité de l'information. Contrairement à d'autres méthodes d'analyse, l'AFC permet de retrouver une structuration naturelle des données, en associant les actions aux éléments de contexte observés. Une telle structuration rend possible la recommandation de ces actions grâce aux règles d'association.

La méthodologie de gestion conceptuelle de contexte proposée a été appliquée dans une étude de cas réalisée à partir des observations de contexte d'un utilisateur réel dans son quotidien.

5. Actions futures

L'approche et la solution proposées permettent d'envisager de nombreuses perspectives. La première consiste à concevoir des mécanismes pour automatiser les interprétations et les recommandations associées. Nous exploiterons pour cela les concepts générés par le treillis de Galois et tous les liens entre ces concepts, afin d'en déduire de manière automatique une interprétation et des recommandations.

Nous avons jusqu'ici utilisé un nombre limité d'éléments de contexte (localisation géographique, créneau horaire, réseau d'accès et mémoire du terminal). Nous étudierons dans l'avenir la pertinence d'autres types d'éléments de contexte afin d'étendre notre approche.

Nous souhaitons également appliquer notre méthodologie à d'autres cas d'étude, et réaliser une évaluation des recommandations effectuées auprès des utilisateurs. À plus long terme, nous souhaitons effectuer une gestion de contexte en temps réel, qui posera des problèmes de passage à l'échelle du fait du volume de données à analyser et aux contraintes temporelles. Une piste possible consiste à utiliser des approches distribuées pour la construction et la mise à jour des treillis de Galois.

Bibliographie

- Agrawal R., Imieliński T., Swami A., Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD Intl. Conference on Management of data, Washington, USA, June 1993, pages 207–216.
- Agrawal R., Srikant R., Fast Algorithms for Mining Association Rules. Proceedings of the 20th International Conference on Very Large Databases, June 1994, p. 478–499.
- Brézillon P., “Using Context for Supporting Users Efficiently”. In: Sprague, R.H. (ed.) 36th Hawaii International Conference on System Sciences (HICSS 2003), IEEE Computer Society Press, Los Alamitos, 2003, pp. CD Rom.
- Chalmers D., Dulay N., Sloman M., Towards Reasoning About Context in the Presence of Uncertainty. 1st international workshop on advanced context modelling, reasoning and management, Nottingham, UK, September 2004, pp. 1-33.
- Chari T., Zouari M., Laforest F., Ontology Based Context-Aware Adaptation Approach. Book Chapter In: Context-Aware Mobile and Ubiquitous Computing for Enhanced Usability: Adaptive Technologies and Applications. Dragan Stojanovic (Ed) ISBN 978-1-60566-290-9. IGI publishing 2009. Avril 2009, pp. 26-58.
- Conan D., Rouvoy R., Seinturier L., “Scalable Processing of Context Information with COSMOS”, Proc. 7th IFIP International Conference on Distributed Applications and Interoperable Systems, Springer-Verlag, Lecture Notes in Computer Science Volume 4531, Paphos, Cyprus, June 2007, pages 210-224.
- Dey A.K., Understanding and using context. Personal and Ubiquitous Computing, vol. 5, n° 1, 2001, pp. 4-7.
- Najar, S., Kirsch Pinheiro, M., Souveyet, C., Mécanisme de prédiction dans un système d'information pervasif et intentionnel. In 8èmes Journées Francophones Mobilité et Ubiquité (UbiComp 2002), (Anglet, France : Cépaduès Editions), p. 146–157.
- Pignotti E., Edwards P., Grimnes G.A., Context-aware personalised service delivery. European Conference on Artificial Intelligence, ECAI 2004, 2004, 1077-1078.
- Priss U., Formal Concept Analysis in Information Science. In: Blaise, C. (ed.) Annual Review of Information Science and Technology, ASIST, vol. 40, 2006, pp. 1-22.
- Schilit, B.N., Theimer, M.M., Disseminating active map information to mobile hosts. IEEE Network, vol. 8, n°5, septembre/octobre 1994, pp. 22-32.
- Sigg S., Haseloff S., David K., An Alignment Approach for Context Prediction Tasks in UbiComp Environments. Pervasive Computing, vol.9, n. 4, 2010, p. 90-97.
- Vanrompay Y., Kirsch-Pinheiro M., Berbers Y., Service Selection with Uncertain Context Information. In Handbook of Research on Service-Oriented Systems and Non-Functional Properties: Future Directions, Reiff- Marganiec S. and Tilly M. (Eds.), IGI Global, 2011, 192-215.
- Wen Q., He J., Personalized Recommendation Services Based on Service-Oriented Architecture. In: IEEE Asia-Pacific Conference on Services Computing, 2006, pp. 356–361.
- Wille R., Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies. Formal Concept Analysis, B.Ganter et al., eds., Springer-Verlag, 2005, pp. 1-33.

Transformation de modèles de lignes de produits en environnement hétérogène

Farah Maamar

*Centre de Recherche en Informatique, Université Paris 1 Panthéon-Sorbonne
90, rue de Tolbiac
75013 Paris*

farah.maamar@malix.univ-paris1.fr

MOTS-CLÉS : Familles de produits, Lignes de produits, interopérabilité, Transformation, l'ingénierie des exigences, le modèle pivot.

KEYWORDS: Software Product Lines, interoperability, Transformations, Requirement engineering, Pivot Model.

ENCADREMENT: Camille Salinesi (PR) et Raúl Mazo peña (MCF)

1. Contexte

Le domaine de l'ingénierie de ligne de produits est en cours de développement, dans un monde où la concurrence est de plus en plus rude et où les entreprises sont de plus en plus amenées à diversifier leurs gammes de produits. Les entreprises les plus compétitives et innovantes se dirigent vers ce nouveau paradigme, afin d'augmenter leurs productivités et diminuer le temps de réalisation du produit. L'identification des points variables et communs d'une ligne de produits dans un modèle est une approche clé dans l'ingénierie des lignes de produits et plus particulièrement pour la gestion d'éléments commun et variable. Le processus de développement d'une ligne de produits passe par plusieurs phases, chaque phase est réalisée par un outil différent, si bien que des incohérences peuvent avoir lieu. Pour assurer la bonne collaboration et la cohérence de ces outils, il est nécessaire de faire face aux problèmes d'interopérabilité de ces outils. Dans cet article nous considérons que la transformation du modèle est l'une des solutions clés d'interopérabilité.

2. État de l'art

Dans la littérature, nous trouvons beaucoup de travaux en rapport avec l'interopérabilité. Nous pensons que dans le cadre de l'ingénierie dirigée par les modèles, la transformation de modèle peut être un moyen pour résoudre le problème

d'interopérabilité. L'IEEE Standard Computer Dictionary définit l'interopérabilité comme étant la capacité de deux ou plusieurs systèmes ou composants de pouvoir échanger de l'information et de pouvoir l'utiliser. Dans le contexte de transformation de modèles, nous soulignons que les systèmes qui nous intéressent sont des outils d'ingénierie de modèles de ligne de produits, et dans ce contexte, une définition précise pour l'interopérabilité a été proposée par (Jouault et al., 2007), l'interopérabilité est la capacité d'exécuter des programmes écrits dans un langage avec des outils conçus pour un autre langage, selon un consensus opérationnel et sémantique, par exemple: exécuter un code ATL avec l'outil QVT-compliant engine.

Dans (Mazo et al., 2011), un travail qui vise à soulever le problème de configuration du modèle de ligne de produits, ils ont pensé à automatiser le processus de raisonnement sur les caractéristiques de modèle. L'approche de transformation qu'ils proposent consiste à appliquer un ensemble de règles de transformation sur 60 modèles (FM) de 2000 caractéristiques, ils ont mis au point pour cela deux stratégies complètement automatisées dans leurs outils VARIAMOS.

(Levendovszky et al., 2002) définit le mapping comme étant un ensemble de règles de transformation de modèles permettant de traduire des instances d'un méta-modèle source en instance d'un méta-modèle cible, le travail de (Baïna et al., 2006) a confirmé qu'il existe une forte corrélation entre l'interopérabilité et le mapping de modèles et a proposé la définition suivante : Soit A et B deux méta-modèles, A et B sont dits interopérables si et seulement s'il existe un mapping bijectif de M_A vers M_B , i.e. qu'il est possible de créer une instance de modèle A à partir de l'instanciation de modèle B, et vice-versa. À partir de cette définition, il a classé trois niveaux de mapping et par conséquent trois niveaux d'interopérabilité.

3. Problématique

Il existe plusieurs outils de modélisation de MLPs et d'autres outils pour les analyser et les configurer, chaque outil a son propre formalisme dont lequel il présente le MPLs, après la modélisation, il est nécessaire d'analyser et configurer le MLPs, i.e. la gestion d'éléments commun et variable. C'est pour cette raison que c'est important qu'un formalisme de modèle donné soit convertible en d'autres formalismes utilisés par les outils d'analyses, i.e. comme les solveurs connus tels que GnuProlog, SAT, BDD et bien d'autres. Dans ce contexte de transformation de modèles, nous allons préparer un environnement propice pour la collaboration et l'interopérabilité entre ces différents outils d'ingénierie de ligne de produits. Certaines questions d'interopérabilité nécessitent l'étude et la vérification:

- Comment assurer l'interopérabilité entre les outils de modélisation et les outils de configurations ?

Il existe quelques approches d'interopérabilité dans le contexte de transformation de modèle qui sont citées dans la section d'état de l'art, et d'autres comme l'utilisation d'annotations sémantiques basées sur les ontologies, ce qui nous amène à poser la question suivante :

- Quelle démarche est la plus adéquate pour assurer l'interopérabilité dans le contexte de transformation des modèles ?

4. Actions réalisées

Dans le cadre de notre travail, il est difficile de se positionner dans les niveaux définis par (Baïna et al., 2006) dans la section 2, car d'un côté nous avons traité une transformation dans un seul sens et de l'autre côté tout concept du modèle \mathbf{M}_a a son équivalent dans le modèle \mathbf{M}_b , mais pas en isomorphisme, en effet, le concept du modèle pivot que nous employons pour la transformation, est un modèle général, qui doit contenir tous les éléments de modèle cible donc chaque concept du modèle source doit avoir son équivalent dans le modèle cible. Dans le contexte de notre travail, «Le modèle Pivot» est un méta-modèle du modèle cible, ce méta-modèle doit être le plus complet possible en contenant tous les éléments nécessaires de modèle cible. Pour but d'obtenir la description la plus exhaustive possible. Le modèle source qui va être transformé, doit être conforme au méta-modèle (modèle Pivot). Ceci implique donc la création d'une couche d'abstraction supplémentaire qui se traduit par l'ajout d'une autre étape de transformation en l'occurrence celle du modèle pivot vers le modèle cible. L'intérêt de ce modèle réside dans son indépendance de tout langage de modèle source. Le modèle pivot est le modèle qui est au centre de la transformation, de telle sorte que ce dernier permet de transformer tout formalisme du modèle source en modèle cible décrit dans le modèle pivot (méta-modèle).

Nous avons présenté dans cet article une solution d'interopérabilité par transformation de MLPs en utilisant le modèle Pivot, le cadre applicatif de notre approche de transformation de modèle peut être défini mathématiquement de la façon suivante :

$$(1) \begin{cases} (F_1 (MS1) = X1 \wedge P_{mmc} (X1))=MC \\ (F_2 (MS2) = X2 \wedge P_{mmc} (X2))=MC \\ (F_3 (MS3) = X3 \wedge P_{mmc} (X3))=MC \end{cases} \Leftrightarrow (F_i(MS_i)=X_i \wedge P_{mmc}(X_i))=MC, i/i \in \mathbf{N}$$

Nous illustrons le modèle (1) par un schéma qui décrit la relation entre les fonctions et leurs enchainements comme suit :

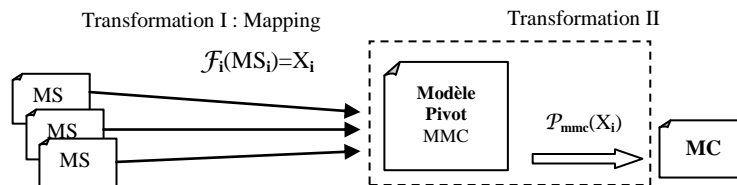


Figure 1 : processus de transformation de MLPs

Les différents formalismes de modèle source sont notés $MS_{i/i \in \mathbb{N}}$, les différents types de modèle cible du même formalisme sont notés MC, les règles de première transformation sont notées F, les différents formalismes de MLPs sont notés $i/i \in \mathbb{N}$, le méta-modèle de modèle cible est noté mmc.

L'approche de transformation est constituée de deux étapes de transformation : une transformation du modèle source en modèle Pivot (transformation I) et du modèle Pivot vers le modèle cible (transformation II). L'application du modèle Pivot nécessite une seule transformation $\mathcal{P}_{mmc}(X_i)$ en modèle cible MC et donc un seul ensemble de règles de transformations applicables pour toute transformation de $MS_{i/i \in \mathbb{N}}$, car le modèle Pivot joue le rôle du modèle intermédiaire vers le même formalisme de modèle cible MC. En effet, la nécessité d'avoir une interopérabilité plus flexible et générale nous a encouragés à penser à une solution plus ouverte,

5. Actions futures

Dans de futurs travaux, nous allons essayer de faire face à la même problématique de l'interopérabilité mais au niveau sémantique, pour cela, i) nous allons tester le mapping bijectif, ii) Nous allons proposer aussi un mécanisme pour détecter la perte d'information qu'elle soit sémantique ou syntaxique. Il est évident que le modèle pivot ainsi défini, ne permet pas de couvrir totalement la sémantique de chaque modèle. Donc il est nécessaire d'enrichir ce modèle pivot par d'autres formalismes de modèle cible. Tester l'application de l'approche proposée sur des différents formalismes de MLPs, ça nous permettra d'améliorer notre approche et façonnera les décisions que nous aurons à prendre.

Bibliographie

- Mazo R., Salinesi C., Diaz D., Alberto L-M., " Transforming attribute and clone enabled feature modèles into constraint programs over finit domains ", September 2011, ENASE 2011: Beijing, China, p.188-199.
- Levendovszky T., Karsai G., Maroti M., Ledeczi A., Charaf H., " Model Reuse with Metamodel-Based Transformations ", in Proceedings of the 7th International Conference on Software Reuse: Methods, Techniques , and Tools, p. 166 -178, 2002.
- Baïna S., interopérabilité dirigé par les model, thèse, Université Henri Poincaré, Nancy I, 2006.
- Jouault F., Kurtev I., "On the interoperability of model-to-model transformation languages", ATLAS Group, INRIA and LINA, University of Nantes, France, 14 May 2007.

Analyse et modélisation de la participation et du E-learning au moyen d'outils collaboratifs

Guillaume Blot

*Laboratoire SND: Sciences, Normes, Décision
Université Paris-Sorbonne & CNRS - FRE 3593
28 rue Serpente
75006 PARIS*

guillaume.blot@paris-sorbonne.fr

MOTS-CLÉS : Participation, E-learning, Média sociaux.

KEYWORDS: Participation, E-learning, Social Media.

ENCADREMENT: Daniel Andler (PR), Pierre Saurel (MCF) & Francis Rousseaux (PR)

1. Contexte

L'objectif de cette thèse est de mener à bien une démarche originale autour des concepts de participation et d'apprentissage afin de définir un outil technique en adéquation avec les multiples facettes que présentent ces deux concepts. La méthode consiste à simultanément, d'une part analyser les concepts de participation et d'e-learning de manière théorique et d'autre part développer le dispositif technique. Ainsi, nous mettons en œuvre deux dynamiques de recherche : l'une définit les usages, l'autre propose des modèles et un socle technique pour mesurer les impacts.

On dispose de deux plateformes d'expérimentation : Moodle, logiciel LMS (Learning Management System) open-source, et Openrendezvous, un agenda partagé, basé sur les dernières technologies Web¹.

1. Guillaume Blot est le concepteur et le développeur de ce système. Actuellement disponible en V1.0 : www.openrendezvous.com.

2. État de l'art

La participation est un concept au cœur de l'utilisation actuelle des technologies numériques. Aucun système ne peut revendiquer une appartenance au paradigme Web 2.0, sans mettre en œuvre des mécanismes participatifs. Contrairement au modèle *top-down*, où l'utilisateur est un simple récepteur, le modèle participatif fait de l'utilisateur un créateur de contenu. Cette participation se caractérise par les inscriptions numériques produites (Flon et al., 2009). Les médias sociaux existent grâce à ces inscriptions. Ainsi, il est possible de créer des structures de données spécifiques, d'optimiser les méthodes d'analyse et d'adapter les algorithmes aux comportements des usagers.

De là, nous pouvons lever une ambiguïté en dissociant de ce concept transversal de participation, certains domaines de recherche comme par exemple, la participation publique, qui vise à administrer la participation citoyenne aux débats publics et aux processus de décision (Blondiaux et al., 2011). Rendue possible à une nouvelle échelle grâce aux médias sociaux, la démocratie participative fédère une communauté pluridisciplinaire de professionnels. La question n'est pas de savoir si ces technologies vont modifier la participation citoyenne, mais comment cette participation va s'organiser dans un environnement médiatique inédit (Rousseaux et al., 2012 ; Monnoyer-Smith, 2011) ?

Les interrogations sont les mêmes à propos du E-learning. Les plateformes LMS (Learning Management System), les blogs, les wiki, les e-portfolios sont autant d'exemples qui démontrent une réorganisation, ayant pour résultat de placer les apprenants au centre du processus d'apprentissage. Ces initiatives prennent une nouvelle forme avec l'apparition des MOOC (Massively Open Online Course) et des cMOOC, modules d'apprentissages à grande échelle, connectivistes, modulaires et ouverts à tout public (Yeager et al., 2013).

3. Problématique

L'apprentissage et la participation citoyenne en ligne partagent un socle commun de pratiques. Pour améliorer les usages dans ces deux contextes, il est important de travailler sur les axes suivants : la liberté d'action, la communication des informations, la transparence, la concertation et la mise en relation. Comment organiser ces mécanismes grâce aux technologies sociales ?

Tout ceci s'organise autour de contraintes organisationnelles : les référentiels de formation pour l'apprentissage et le droit pour la participation citoyenne. Ces innovations se substituent-elles aux procédés traditionnels ou viennent-ils s'y greffer ? Comment les y intégrer ?

Il apparaît que ces deux concepts sont clivés dans une configuration identique. D'un côté, certains voient en ces pratiques une hyper-sociabilité, l'amélioration de l'esprit critique et la mise en œuvre d'appropriations personnelles (Le Boterf, 2013 ; Jenkins, 2006). D'un autre côté, les plus sceptiques dénoncent des inégalités d'accès

aux services, la désagrégation des communautés, la fragmentation de la société et le manque de mesures de validation (Putnam, 2000).

L'objectif de ce travail est de construire des modèles, de produire des outils et de mettre en lumière leurs pertinences, à partir de l'analyse d'une part des moyens techniques et d'autre part des usages et des contraintes organisationnelles.

4. Actions réalisées

4.1. *Time graph : conception d'un Réseau social basé sur le temps*

La mise en relation d'usagers à travers un système est organisée à partir de structures de données spécifiques. Pour être qualifié de réseau social, les interactions doivent s'organiser à l'intérieur d'un graphe, ayant une certaine topologie : il faut éviter d'avoir plus de deux classes de nœuds et les poids des arcs doivent être du même ordre (Wasserman, 1994). De plus, d'autres mesures sont effectuées pour disposer d'une analyse plus précise (Barrat, 2004).

L'agenda partagé Openrendezvous, permet à des personnes de prendre rendez-vous à distance avec des prestataires (médecin, salon de coiffure, garagistes, administration. . .). Dans ce cadre, la modélisation d'un *Time graph* a été proposée afin d'identifier des séquences de rendez-vous et de prédire les disponibilités des usagers. Ici, les nœuds sont des événements et les arcs directionnels sont valués grâce à un coefficient relatif à la durée qui sépare deux nœuds.

4.2. *Le design pattern MapReduce*

Grâce à la philosophie du *design pattern* MapReduce, hérité des langages de programmation fonctionnels, il est possible de déployer un système à inscriptions numériques, sur une très grande échelle (Dean et al. 2004). Sur le modèle des MOOC, conçus pour accueillir un nombre conséquent d'utilisateurs, un modèle de données basé sur XML a été réalisé. Grâce à un accès fragmenté aux données, on récupère instantanément l'environnement d'un utilisateur sans avoir à parcourir toute la base. En effet, la topologie des relations d'un système relationnel peut impliquer la surcharge des tables de jointures. Soit le schéma relationnel suivant, relatif au *use-case* Openrendezvous :

ETABLISSEMENT(#idet) - CLIENT(#idcli) - RDV(#idet,#idcli)

Pour le modèle relationnel, on constate une progression linéaire du temps d'accès, en fonction du nombre d'établissements présents dans la base. Avec 1000 établissements présents et 100 RDV par établissement en moyenne, la récupération de la base se fait en 0.03sec et avec 15000 établissements, le temps de réponse est 0.8sec. En revanche le temps d'accès est constant avec le modèle fragmenté.

5. Actions futures

La constitution de structures d'analyse constitue l'axe majeur de recherche pour la suite des travaux. En plus d'optimiser le *Time graph*, il est indispensable de trouver des procédés permettant d'y intégrer d'autres mesures que le temps, en associant, par exemple, l'événement, l'utilisateur, le lieu, la durée et le type de relation au sein de la même structure. Pour avancer sur cet axe il est prévu d'organiser un environnement multi-agents disposant d'outils de visualisation des interactions. Deux types de démarches seront alors envisageables. La première consiste à lancer différentes procédures multi-agents sur un même objet. La seconde consiste à étudier un même groupe de règles sur des objets différents.

Bibliographie

- Barrat A., Barthélemy M., Pastor-Satorras R., Vespignani A., *"The architecture of complex weighted networks"*, PNAS journal, vol. 101 no. 11, p 3747-3752, 2004.
- Blondiaux L., Fourniau J-M., *"Un bilan des recherches sur la participation du public en démocratie : beaucoup de bruit pour rien ?"*, Participations, vol.1, p. 10-34, 2011.
- Dean J., Ghemawat S., *"MapReduce : Simplified data processing on large clusters"*, Google Inc : OSDI, 2004.
- Flon E., Davallon J., Tardy C., Jeanneret Y., *"Traces d'écriture, traces de pratiques, traces d'identités"*, H2PTM'09, Paris, Editions Hermès-Lavoisier, p 181-192, 2009.
- Jenkins H., *Convergence culture : Where old and new media collide*, New York, New York University Press, 2006.
- Le Boterf G., *Travailler en réseau et en partenariat. Comment en faire une compétence collective ?*, ED 3, Editions Eyrolles, 2013.
- Monnoyer-Smith L., *"La participation en ligne, révélateur d'une évolution des pratiques politiques ?"*, Participations, vol. 1, 2011, p. 156-184.
- Putnam R., *Bowling Alone : The collapse and revival of American community*, New York, Simon & Schuster, 2000.
- Rousseaux F., Soulier E., Saurel P., Neffati H., *"Agencement multi-échelle de territoires à valeur ajoutée numérique : des Hétérotopies foucaaldiennes aux Complexes simpliciaux"*, Politiques publiques & Systèmes complexes, Editions Hermann, 2012, p. 169-193.
- Wasserman S., *Social Network Analysis : Methods and Applications*, Cambridge University Press, 1994.
- Yeager C., Hurley-Dasgupta B., Bliss C., *"CMOOCs and Global learning : An authentic alternative"*, JALN, vol. 17, 2013, p. 133.

Conception d'un système expert à base de connaissances floues d'aide à la décision pour la maladie d'Alzheimer

Firas Zekri

*Faculté des Sciences Economiques et de Gestion de Sfax, Laboratoire MIR@CL
Route de l'Aéroport km 4
B.P. 1088, 3018 Sfax, Tunisie
firas3112@gmail.com*

MOTS-CLÉS: Maladie d'Alzheimer, Système d'aide à la décision, Système expert, logique floue, système multi-agents, ontologie floue.

KEYWORDS: Alzheimer Disease, Decision support system, Expert system, Fuzzy logic, multi-agent system, Fuzzy ontology

ENCADREMENT: Rafik Bouaziz (PR)

1. Contexte

La Maladie d'Alzheimer (MA) est la plus fréquente des maladies neuro-dégénératives, c'est-à-dire les maladies liées à une détérioration irréversible du fonctionnement des cellules du système nerveux. La lutte contre la MA est devenue un enjeu majeur aussi bien dans les pays développés que dans les pays en cours de développement, où le progrès de la médecine a prolongé la durée de vie de la population. En effet, l'informatique a su apporter une aide appréciable pour la prise de décision dans divers domaines, y compris celui de la médecine. Nos travaux se situent effectivement dans le cadre des recherches en ingénierie et management des connaissances médicales. Ils sont principalement organisés autour des problèmes relatifs à la conception des systèmes d'aide à la décision pour la MA.

2. État de l'art

Les systèmes d'aide à la décision médicale ont fait l'objet de multiples réalisations. À notre connaissance, les systèmes experts sont les plus utilisés à cet effet. Un système expert est un produit de l'intelligence artificielle, qui permet de simuler le raisonnement d'un expert. Actuellement, divers travaux traitent de la réalisation de systèmes experts pour le diagnostic et la détection de la MA. *Knowledge Based Clinical Decision Support System (KBCDSS)* est un système d'aide à la décision pour la détection précoce de la MA (Sanchez *et al.*, 2011). Il a

pour but d'aider les médecins dans les premiers stades de la MA. Il se base sur l'utilisation des connaissances pluridisciplinaires rassemblées, et sur l'inférence et le raisonnement fondés sur des bases de connaissances sous-jacentes. Les auteurs de (Obi *et al.*, 2011) ont élaboré un système flou de réseaux de neurones pour le diagnostic de la MA. Dans ce travail, la procédure traditionnelle employée par le médecin pour le diagnostic médical de la MA est analysée en utilisant une procédure d'inférence « neuro-flou ». Le système proposé est basé sur une approche d'aide à la décision appropriée au diagnostic de la MA. L'avantage majeur de ce système est sa capacité de manipuler des informations imparfaites et imprécises. Cependant, les résultats donnés sont très réduits. Ils sont limités à trois réponses : Alzheimer, peut-être Alzheimer ou pas d'Alzheimer. Les auteurs de (Martinez-Murcia *et al.*, 2012) et (Savio *et al.*, 2012) ont démontré que le diagnostic assisté par ordinateur permet aux médecins de détecter les premiers stades de la MA, et que les images fonctionnelles du cerveau se révèlent être très utiles dans cette tâche. NES (Neurologist Expert System) est un système expert à base de règles qui utilise une séquence interactive de questions-réponses pour le diagnostic des maladies neurologiques élaboré par (Al-Hajji, 2012). Ce système est destiné à être utilisé comme un système de consultation par les neurologues et chercheurs afin de prendre une décision. Cependant la génération de dix types de cas de troubles neurologiques dans ce système affaiblit le diagnostic et les résultats de chaque maladie. De plus, les thérapies données en résultats sont des thérapies générales et ne sont pas spécifiques au cas de la MA.

3. Problématique

Les études sur la MA ont montré que la lutte contre cette maladie dépasse l'ensemble des consultations chez un médecin. Cette lutte exige un grand effort de plusieurs acteurs qui doivent tous participer à faire réussir le processus de soin du patient. En effet, la prise en charge médicale de la MA doit faire appel à une équipe pluridisciplinaire (neurologues, psychiatres, gériatres, orthophonistes, etc.), et elle doit s'adresser aussi bien aux patients qu'aux personnes considérées comme des « aidants » (infirmières, aide-ménagères, conjoints, membres de famille). En outre les études ont montré que cette maladie s'exprime différemment selon les personnes. Chaque malade présente des symptômes qui lui sont propres : certains signes peuvent apparaître fortement dès le début de la maladie chez un patient, ou à l'inverse, rester très discrets même au bout de plusieurs années chez un autre patient.

La problématique à traiter dans ce projet de thèse consiste à voir comment on peut réunir et diriger tous ces efforts pour prendre la bonne décision pour le bien du patient. Comment on peut soulager le travail des aidants pendant leurs journées avec le patient et leur permettre de maintenir une activité professionnelle ? Comment bénéficier de toutes les connaissances provenant des experts du domaine, des aidants et des proches du patient ? Comment être en mesure de répondre rapidement aux appels d'aide à la décision ?

Nous visons alors à définir une approche pour le développement d'outils informatiques performants qui permettent d'aider les médecins dans le diagnostic et dans les prises de décision concernant les traitements à faire, et d'assister les aidants dans leur travail ainsi que les proches du patient dans leur soutien, en assurant une bonne gestion des informations et des connaissances qui peuvent contribuer à un bon diagnostic et un traitement efficace du patient.

4. Actions réalisées

Nous avons premièrement étudié la maladie d'Alzheimer. Ensuite nous avons effectué une recherche bibliographique approfondie sur le domaine des systèmes d'aide à la décision médicale, et plus précisément la construction et l'utilisation des systèmes experts et leur rôle dans l'aide à la décision médicale. En complément, nous avons étudié un autre domaine en intersection, qui est la logique floue, et leur apport pour le calcul dans le moteur d'inférence d'un système expert. Puis nous avons essayé de satisfaire les exigences relevées dans la problématique en proposant un processus d'aide à la décision pour la MA. Ce processus contient trois étapes qui sont : étape de collecte de données, étape d'analyse et inférence et étape de rappel automatique (cf. figure 1). Ce processus est basé sur l'intervention avec les experts du domaine d'une part, et sur l'intervention avec l'expert responsable au cas pathologique d'une autre part. La première intervention consiste à alimenter le système par des connaissances expertes et la deuxième consiste à valider les résultats trouvés pour garantir leur utilisation par la suite. Ces résultats sont relatifs à la quantité et la qualité des données transmises par les différents acteurs du système.

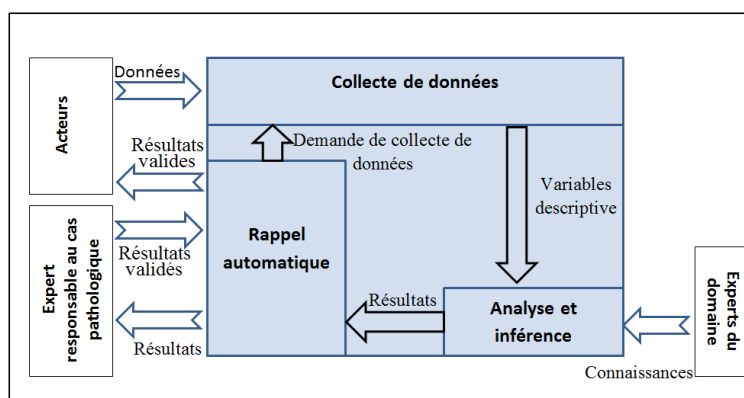


Figure 1 : Processus d'aide à la décision pour la MA

Nous avons proposé par la suite une architecture pour un système expert à intégrer dans l'étape « analyse et inférence » (cf. figure 2). Cette architecture est basée sur la spécialisation floue des connaissances où chaque patient a sa propre base de connaissances. Elle est conçue pour combler le problème d'hétérogénéité dans la MA et le problème de temps d'inférence. En résolvant ces deux problèmes, le système sera capable de générer des résultats plus efficaces et plus rapides.

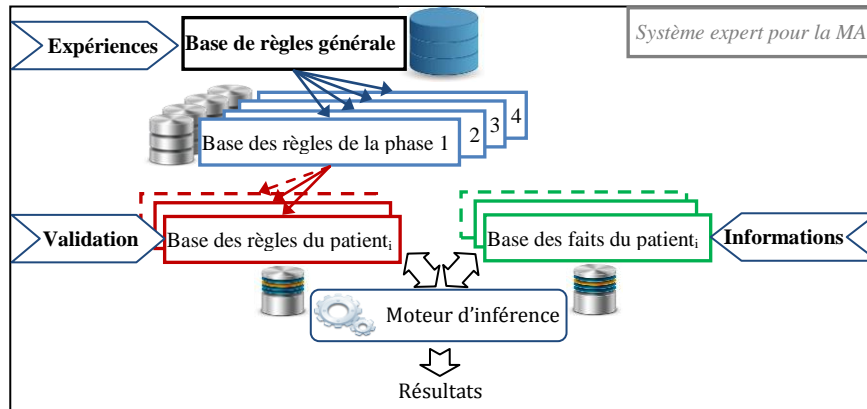


Figure 2 : Architecture d'un système expert pour la MA

5. Actions futures

Dans les prochaines étapes, nous allons explorer et proposer des voies de solution aux autres problèmes qu'on peut rencontrer, comme l'acquisition des connaissances floues, le raffinement des connaissances floues et l'inférence floue. Ensuite, nous souhaitons définir une ontologie floue que nous allons appeler « Alzheimer-OntoF » permettant de résoudre les problèmes sémantiques relatifs aux informations imparfaites et imprécises, et les problèmes de communication. Nous comptons définir une approche multi-agents basée sur le processus d'aide à la décision proposé pour assurer plus d'autonomie. Enfin, nous terminons par développer et opérationnaliser un système d'aide à la décision pour la MA, en tant que prototype mettant en œuvre les contributions que nous visons.

Bibliographie

- Al-Hajji A., "Rule-Based Expert System for Diagnosis and Symptom of Neurological Disorders -Neurologist Expert System (NES)-", *International Conference on Communications and Information Technology ICCIT*, 2012, p. 67-72.
- Martinez-Murcia F. J., Gorriz J. M., Ramirez J., Puntonet C. G., Salas-Gonzalez D., "Computer Aided Diagnosis tool for Alzheimer's disease based on Mann-Whitney-Wilcoxon U-Test", *Expert Systems with Applications*, Vol.39, 2012, p. 9676-9685.
- Obi J.C., Imainvan A. A., "Decision Support System for the Intelligent Identification of Alzheimer using Neuro Fuzzy logic", *International Journal on Soft Computing (IJSC)*, Vol.2, No.2, May 2011, p. 25-38.
- Sanchez E., Toro C., Carrasco E., Bonachela P., Parra C., Bueno G., Guijarro F., "A Knowledge-based Clinical Decision Support System for the diagnosis of Alzheimer Disease", *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, 2011, p. 351-357.
- Savio A., Graña M., "Deformation based feature selection for Computer Aided Diagnosis of Alzheimer's Disease", *Expert Systems with Applications*, Vol.40, 2013, p. 1619-1628.

Vers un processus d'amélioration continue pour la capitalisation des connaissances à *STMicroelectronics*

Manel Brichni

LIG, 220 Rue de la Chimie, BP 53, 38041 Grenoble, France

GSCOP, 46 avenue Félix Viallet, 38031 Grenoble Cedex 1, France

manel.brichni@g-scop.grenoble-inp.fr

MOTS-CLÉS : Capitalisation des connaissances, Wiki, approche centrée utilisateur

KEYWORDS: Knowledge capitalization, Wiki, user centred approach

ENCADREMENT. Sophie Dupuy-Chessa (LIG), Lilia Gzara (GSCOP) et Corinne Jeannet (STMicroelectronics)

1. Contexte

La croissance rapide des entreprises, le départ des employés, la complexité de la nouvelle technologie et la prolifération rapide de l'information, sont des raisons pour lesquelles les entreprises cherchent à tirer profit de leur expertise (Bera, 2002; Lyn, 2009). La capitalisation des connaissances offre aux utilisateurs un moyen de stocker et de réutiliser les expériences accumulées au cours de leur travail quotidien, en moins de temps (Gardoni *et al.*, 2005). Le processus de capitalisation comprend plusieurs étapes (Bera, 2002; Schreiber, 2000) qui chacune soulève des questions. À *STMicroelectronics*, la fabrication des semi-conducteurs est un processus complexe qui doit être constamment contrôlé et amélioré. En particulier, l'équipe du *Reporting* est confrontée à des difficultés dans la gestion de ses rapports dont le nombre est en forte croissance, tandis que la connaissance sur leur création est de plus en plus perdue. Ainsi, non seulement une mauvaise gestion des connaissances explicites mais aussi une perte des connaissances tacites sont soulignées. Par conséquent, nous appliquons

notre approche sur le processus du *Reporting*, pour laquelle l'équipe rencontre des problèmes de capitalisation de connaissances.

2. État de l'art

La connaissance doit être capitalisée de manière simple. Le processus de capitalisation est composé des trois étapes suivantes :

- L'identification des connaissances, qui consiste à extraire et identifier les connaissances qui doivent être capitalisées, avec la participation des utilisateurs.
- La représentation des connaissances, qui consiste à formaliser les connaissances identifiées. Dans la littérature, la connaissance peut être représentée selon ses trois facettes le *Pourquoi* (Frey, 2013), le *Comment* (Ouerteni, 2007) et le *Quoi* (Page *et al.*, 2000) de la connaissance.
- Le stockage des connaissances, dont l'objectif est de conserver une trace des problèmes rencontrés et les solutions identifiées pour des éventuelles réutilisations.

Les connaissances sont essentiellement capitalisées à des fins de partage et réutilisation. L'idée est de fournir un moyen simple de partage tenant compte des pratiques, des besoins quotidiens et de la culture de l'entreprise. À *STMicroelectronics*, depuis quelques années, les employés ont opté pour un *Wiki* pour partager leurs connaissances. C'est pour cela, nous avons étudié son utilisation pour assister le travail collaboratif dans plusieurs domaines (Lyn, 2009; Garcia-Perez *et al.*, 2010; Parker *et al.*, 2007; Ruslan *et al.*, 2009). Par ailleurs, faciliter l'accès aux connaissances partagées aide l'utilisateur à les réutiliser d'une manière efficace. Dans la plupart des entreprises, les connaissances évoluent et changent régulièrement. C'est pour cela que la capitalisation des connaissances doit se faire de manière continue en tenant compte de leur évolution. Ceci explique le cycle dans le processus de capitalisation des connaissances.

De nombreuses approches ont discuté le problème de la capitalisation des connaissances. Nous citons, par exemple, l'approche CommonKads (Schreiber, 2000) héritée de l'ingénierie des connaissances. Nous retenons de cette approche, la construction de sa collection de modèles afin de capturer les différents aspects de la connaissance tout en assurant leur cohérence (les modèles UML par exemple). La AKM (Active Knowledge Modeling) (Lillehagen *et al.*, 2009) est une autre approche de capitalisation des connaissances. Parmi d'autres avantages, la AKM permet aux utilisateurs industriels de construire et gérer leurs propres environnements de travail. Être en mesure de soutenir le travail collaboratif et l'apprentissage participatif dans la gestion des connaissances et la manière dont elles peuvent être représentées, déterminera la qualité de l'AKM.

3. Problématique

Ma thèse consiste à proposer un processus d'amélioration continue pour la capitalisation des connaissances, considérant simultanément la participation des utilisateurs et l'intégration de la technologie. Cependant, le principal défi consiste à associer le

contexte industriel, les systèmes existants, les pratiques organisationnelles et les éventuelles évolutions. Nous estimons que la participation des utilisateurs est un élément important dans le processus de capitalisation des connaissances. C'est le but de notre approche centrée utilisateur. Pour faire face à cette problématique, certaines étapes cruciales doivent être réalisées. Pour résumer, ce travail de recherche doit répondre aux trois questions suivantes :

- Comment identifier les connaissances existantes, embarquées non seulement dans les processus, mais aussi dans les pratiques organisationnelles ?
- Comment représenter les connaissances identifiées et assurer un partage efficace ?
- Comment assurer un stockage et un partage des connaissances à *STMicroelectronics* favorisant leur réutilisation et leur évolution ?

4. Actions réalisées

4.1. L'évaluation de l'utilisation de *Stiki* pour le partage des connaissances

À *STMicroelectronics*, un *Wiki* a été adopté pour favoriser le partage de connaissances, appelé *Stiki*. Il a été conçu et mis en œuvre en 2009 pour résoudre des problèmes tels que le manque de partage des connaissances, l'existence de différents référentiels de stockage, la recherche efficace d'information, etc. Notre méthodologie d'évaluation est basée sur une approche centrée utilisateur permettant d'identifier les besoins et les attentes utilisateurs et les axes de progrès pour que le *Wiki* accomplisse son objectif. Par ailleurs, nous avons intégré des méthodes d'évaluation quantitatives et qualitatives. La méthodologie d'évaluation proposée est constituée des étapes suivantes : *Stiki Review (des réunions d'observation), définition des hypothèses, analyses statistiques, interviews semi structurées, questionnaire, validation des hypothèses*. Nous avons pu identifier des axes de progrès autour de *Stiki* proposés par les utilisateurs. Par exemple, une formation régulière peut aider à apprendre à l'utiliser et résoudre les problèmes techniques avec son utilisation. De même, fournir un formalisme commun en structurant le contenu des pages facilite l'accès à l'information et guide les utilisateurs dans leurs contributions. Une utilisation régulière assure la pertinence de son contenu et évite son obsolescence. L'intégration de la documentation dans *Stiki* dans le cycle de vie des projets est également un moyen encourageant son utilisation régulière. D'un autre point de vue, nous considérons que la motivation des utilisateurs est une étape cruciale pour les encourager à prendre part au *Wiki*. Pour cela, une des pistes, proposée par plusieurs utilisateurs, consiste à mettre en place un système de notation des pages pour prouver au contributeur que sa page a été utile ou si elle nécessite une révision.

4.2. La représentation des connaissances

Afin d'appuyer le processus de capitalisation, nous cherchons à définir un format pour représenter les connaissances en étudiant ses différentes caractéristiques. En effet,

comme nous avons discuté dans la section 2, nous considérons que la connaissance possède trois caractéristiques différentes et complémentaires : le *Quoi ?* le *Pourquoi ?* et le *Comment ?*. En suivant une approche centrée utilisateur, nous avons pu définir les informations décrivant ces 3 caractéristiques. Elles vont nous permettre de définir les formats des templates pour *Stiki*, offrant ainsi un cadre de capitalisation répondant aux besoins de ses utilisateurs et où chacune décrit un aspect différent de la connaissance. Par conséquent, notre objectif est de capitaliser efficacement à travers *Stiki* les connaissances liées à la création de rapports, tout en fournissant une représentation complète de ses caractéristiques.

5. Actions futures

Dans la suite, nous examinerons le reste des phases du processus de capitalisation. Tout en suivant une approche centrée utilisateur, nous allons étudier comment les connaissances capitalisées pourraient être réutilisées de manière efficace par les utilisateurs. Nous proposerons donc un cycle d'amélioration continue des connaissances afin d'assurer leur évolution par les utilisateurs.

6. Bibliographie

- Bera P., Knowledge requirements analysis : A case study, PhD thesis, Utkal University, India, 2002.
- Frey A. G., Quality of Human-Computer Interaction : Self-Explanatory User Interfaces by Model-Driven Engineering, PhD thesis, Université de Grenoble, 2013.
- Garcia-Perez A., Ayres R., « Wikifailure : the Limitations of Technology for Knowledge Sharing », *Electronic Journal of Knowledge Management Volume*, vol. 8, p. 43-52, 2010.
- Gardoni M., Frank C., Vernadat F., « Knowledge capitalisation based on textual and graphical semi-structured and non-structured information : case study in an industrial research centre at EADS », *Computers in Industry*, vol. 56, p. 55-69, 2005.
- Lillehagen F., Krogstie J., *Active knowledge modeling of enterprises*, Springer, 2009.
- Lyn G. T. P., « Wikis as a knowledge management tool », *Journal of Knowledge Management*, vol. 13, p. 64-74, 2009.
- Ouerteni M. Z., DEPNET : une approche support au processus de gestion de conflit basée sur la gestion des dépendances de données de conception, PhD thesis, Université Henri Poincaré, Nancy I, 2007.
- Page M., Gensel J., Capponi C., Bruley C., Genoud P., Ziebelin D., « Representation de connaissances au moyen de classes et d'associations : le système AROM », *LMO*, Hermès, p. 91-106, 2000.
- Parker K. R., Chao J. T., « Wiki as a Teaching Tool », *Interdisciplinary Journal of Knowledge and Learning Objects*, vol. 3, p. 57-72, 2007.
- Ruslan R., Fawei G., « Researching the use of Wiki's to facilitate group work », *Social and Behavioral Sciences*, vol. 1, p. 2620-2626, 2009.
- Schreiber G., *Knowledge engineering and management*, MIT Press, 2000.

Accords de niveau de service et modèle de réputation pour le contrôle d'usage

Elena Jaramillo Rojas

*LIUPPA, Université de Pau et des Pays de l'Adour
371 rue du Ruisseau – BP 201
40004 Mont de Marsan
gloriaelena.jaramillorojas@univ-pau.fr*

MOTS-CLÉS : objectifs de niveau de service, réputation, workflow inter-organisationnel, gestion de la responsabilité.

KEYWORDS: service-level objectives, reputation, inter-organizational workflow, accountability.

ENCADREMENT. Philippe Anioté (PR) et Manuel Munier (MCF)

1. Contexte

La sous-traitance de services est devenue une nouvelle forme d'interaction pour les entreprises en lui permettant de se concentrer sur leur cœur de métier en profitant des avantages des architectures orientées services. La collaboration résultant de cette sous-traitance entraîne la création de services composés et de workflows complexes dans lesquels le client d'un service partage des informations sensibles avec d'autres organisations externes et indépendantes. Dans un tel contexte inter-organisationnel, plusieurs défis restent à être traités par la communauté. Tout d'abord, le partage explicite de données entre les différents partenaires engendre une perte de contrôle sur les données passées au(x) fournisseur(s) de service(s). Deuxièmement, la complexité de la chaîne de processus nécessite de définir des stratégies de gestion de la confiance et de gestion des responsabilités (et des pénalités) en cas de mauvaise exécution des tâches prévues. Troisièmement, la création de workflows dynamiques permet de poursuivre ces interactions inter-organisationnelles malgré des comportements défectueux tout en préservant le niveau de service requis par le client. Finalement, les partenaires du workflow doivent prouver dans leurs comportements qu'ils respectent

les politiques établies, mais agissant comme des boîtes noires ils doivent garder leur autonomie. Dans cet article, nous présentons une nouvelle analyse de la notion d'accord de niveau de service (SLA) appliquée au contrôle d'usage des données. Pour cela, dans la section suivante, nous détaillons les trois parties principales de notre proposition. Nous identifions ensuite les principales contributions de notre travail par rapport à l'état de l'art. La conclusion et les travaux futurs termineront cet article.

2. Problématique et actions réalisées

Les SLA sont des documents signés, consentis et convenus qui engagent à la fois le client et le fournisseur du service sur le respect d'une liste de conditions quant à la fourniture du service. Plusieurs approches ont été proposées pour combler l'écart entre les aspects juridiques et techniques. Le problème persistant consiste à savoir comment définir ce qu'est un niveau de service dans un environnement inter-organisationnel.

Notion de niveau de service. Savoir ce que signifie un accord de niveau de service inter-organisationnel n'est pas une question triviale, car dans ce genre d'environnements bien que les partenaires aient convenu un objectif final commun, chacun d'eux possède néanmoins ses propres objectifs [1]. Suivant notre approche orientée contrôle d'usage, nous définissons le niveau de service inter-organisationnel en termes de respect des conditions d'utilisation imposées par le propriétaire des données et représentées sous la forme d'objectifs de niveau de service (SLO). Nous proposons également la notion de *niveau de service propagé* (PSL) en tant que stratégie nous permettant de lier et de surveiller les SLO entre les différents partenaires impliqués dans le workflow, et plus important encore, d'adapter et de négocier les SLO en cas de non-conformité. Nous proposons à partir de la notion de PSL que la non réalisation d'au moins un SLO par un nœud du workflow génère un effet domino sur la totalité des SLO en déclenchant un processus de négociation entre les partenaires. Nous considérons qu'avant que les données puissent être utilisées par un autre partenaire du workflow, celui-ci doit être informé de l'état du système et négocier les SLO non respectés. Cela implique que l'algorithme qui régit notre approche de PSL a en entrée les objectifs de service et produit en sortie un résultat final (si nécessaire) et la preuve de l'usage qu'il prétend avoir fait des données tout au long du workflow ; le processus est considéré comme dynamique et dépend complètement de la réalisation (ou non) des SLO. En considérant l'interaction inter-organisationnel trois autres composants sont impliqués, à savoir le coordinateur du workflow, le composant de négociation et le composant nommé *Universal, Description, Discovery and Integration* (UDDI). La principale fonction du coordinateur est de distribuer les fonctions aux nœuds du workflow, d'évaluer l'accomplissement des SLO et, en cas d'un comportement incorrect, d'informer le composant de négociation de la situation afin qu'il essaye, dynamiquement, de trouver dans l'UDDI un autre service qui pourrait garantir l'accomplissement des SLO au niveau global.

Niveau de service et gestion de la responsabilité. Comme il a été indiqué, la sortie du processus est un journal qui sert de preuve de l'usage qui a été fait des

données à l'intérieur du workflow. Afin d'utiliser le journal comme preuve il faut qu'il ait les propriétés suivantes : (i) Il ne peut être modifié ou accessible que par des entités autorisées. (ii) Il doit détailler le flux des données complet, incluant la date à laquelle les données ont été utilisées, le but de l'utilisation, l'action effectuée avec les données ainsi que le responsable de l'action. (iii) Il doit garantir que les informations enregistrées dans le journal sont fiables en signant l'entrée du journal.

Quant à la mise en œuvre, l'architecture pourra être centralisée ou distribuée, mais il faudra que la création du journal soit synchronisée avec l'exécution du workflow, et donc également avec le flux de données. Supposons que les données transmises soient modifiées par deux nœuds indépendants suite à un embranchement du workflow. Puisque le coordinateur du workflow connaît la configuration globale du système, il pourrait éviter les mises à jours incorrectes des données, ou les erreurs d'enregistrement ou de fusion dans le journal.

Niveau de service et réputation. Un point à prendre en compte est la façon dont le composant de négociation choisit un nouveau fournisseur de services pour l'intégrer au workflow. Nous utilisons pour cela la notion de réputation. La première fois qu'un client accède aux services d'un fournisseur, le client lui fait confiance quant à l'accomplissement des SLO convenus. Cependant, en cas de violation, le système applique une pénalité au fournisseur. Au niveau de l'architecture, la base de données contenant les informations sur la réputation de chaque entité ne pourra être accédée que par le composant de négociation afin de préserver l'intimité de l'entité, mais aussi parce que nous proposons une mesure subjective de la réputation basée sur des seuils et les sanctions établies par le client du service en tenant compte seulement de ses objectifs. Dans notre proposition nous attribuons des sanctions pour la violation de SLO liés à l'utilisation des données, mais aussi de la politique de sécurité.

3. État de l'art

Des initiatives de sécurité inter-organisationnelle comme TOrBAC [4] et Multi-Trust-OrBAC [2] incluent dans leur politique de sécurité, comme dans le présent article, un modèle de réputation. Cependant, ces travaux ne comprennent pas la notion de contrôle d'usage des données, ni n'envisage de façon explicite le processus de négociation entre les différents partenaires impliqués dans le workflow. D'autre part, dans la littérature, plusieurs propositions abordent le problème de l'utilisation des données par un troisième partenaire. [3] propose une approche de contrôle de l'usage en fonction de conditions contextuelles qui sont vérifiées pendant et après que l'autorisation ait été accordée. D'une manière générale, à notre connaissance, le contrôle d'usage a été défini et mis en œuvre comme un accès en continu [5]. En tenant compte de l'état actuel de l'évolution de nos domaines de recherche, voici les principales contributions de notre travail : (i) Notre proposition vise à compléter les techniques d'autorisation et d'authentification traditionnelles en passant de *qui* peut accéder à *quelles* ressources et *quand*, à *comment* les ressources sont utilisées une fois accessibles. (ii) Nous proposons une nouvelle approche de l'utilisation des SLO dans laquelle nous démontrons le

phénomène de niveau de service propagé. (iii) Nous mettons en œuvre notre notion de niveau de service propagé pour créer des workflows dynamiques et pour déterminer un degré de réputation pour chaque partenaire.

4. Conclusion

Comment faire pour contrôler les données qui sont partagées ou fournies à un prestataire externe est un domaine de recherche actif dans le monde entier en raison des implications juridiques et commerciales. Dans cet article nous présentons notre proposition de contrôle d'usage des données sur la base de relations contractuelles entre les entités manipulant ces données. Nous visons à établir une chaîne de responsabilités dans l'exécution d'un workflow en créant des journaux, mais aussi avec l'application d'une stratégie de sanction sur la base de la violation de la politique de sécurité et des objectifs de niveau de service convenus quant à l'utilisation des données. Nous proposons aussi l'applicabilité des niveaux de service propagés dans un workflow inter-organisationnel.

5. Actions futures

La prochaine étape consiste à intégrer et à formaliser les SLO en tant qu'élément de la politique de sécurité inter-organisationnelle. Notre idée initiale est orientée vers l'utilisation d'OrBAC ou d'une de ses variantes telles que Multi-Trust-OrBAC [2] et inclu de nouveaux éléments liés au but de l'utilisation des données, en proposant également une nouvelle catégorie pour les conditions contextuelles associées à l'utilisation des données dans un environnement inter-organisationnel.

Bibliographie

- [1] Munier M., Lalanne V., Ardoy P-Y., Ricarde, M., Legal Issues about Metadata Data Privacy vs Information Security, Proc. 8th DPM'2013. LNCS 8247, p. 162-177,
- [2] Ben Saidi, M., Marzouk A., Multi-Trust-OrBAC : Access Control Model for Multi-Organizational Critical Systems Migrated to the Cloud, IJSCE, 2013. p. 51-55
- [3] Sans, T., Cuppens, F., and Cuppens-Boulahia, N., A Framework to Enforce Access Control, Usage Control and Obligations. Annales Des Télécommunications Vol. 62, Issue 11-12 , p. 1329-1352. 2007.
- [4] El Kalam A.A., Marzouk A., TOBAC : A Trust Organization Based Access Control Model for Cloud Computing Systems, IJSCE, 2012 Vol. 2, Issue-4, p. 122 - 130
- [5] Pretschner, A., Schütz, F., Schaefer, C. Policy Evolution in Distributed Usage Control. ENTCS. 2009. Vol. 244, p. 109-123

Application de l'IDM pour la modernisation et la migration d'application existante

Khalil Zouaoui

*Université Lille1, Laboratoire d'Informatique Fondamentale (LIFL)
USTL - Bâtiment M3 extension Cité Scientifique - 59655 - Villeneuve d'Ascq Cedex
khalil.zouaoui@etudiant.univ-lille1.fr*

MOTS-CLÉS: rétro conception, réingénierie, Ingénierie Dirigée par les Modèles.

KEYWORDS: reverse engineering, reengineering, clustering, Model Driven Engineering.

ENCADREMENT: Sami Meftali(PR) et Cédric Dumoulin(MCF)

1. Contexte

La société CDI Technologies est un éditeur de logiciels dédiés à la topographie et à l'aménagement d'infrastructure urbaine. Ces logiciels se composent notamment de modules de dessin assisté par ordinateur (DAO) en 2 et 3 dimensions. CDI Technologies mène un plan de recherche et développement ayant plusieurs objectifs : amélioration de la qualité et des fonctionnalités des logiciels existants ; interopérabilité des logiciels futurs avec les plateformes graphiques du marché (Autocad, Microstation et Polaris) ; développement de nouveaux produits s'inscrivant dans son domaine d'intervention.

Notre sujet de recherche s'inscrit dans le domaine de la réingénierie d'applications existantes avec un double objectif : il s'agit d'une part, pour l'entreprise, de fournir un nouveau produit à partir de l'existant afin de permettre l'intégration de nouvelles exigences et assurant la facilité de maintenance; il s'agit d'autre part, du point de vue scientifique, de formaliser la méthodologie permettant de répondre à des besoins de modernisation d'applications existantes.

Nous présentons dans cet article la méthodologie proposé : nous commençons par un rapide état de l'art, nous posons ensuite en section 3) les problématiques rencontrés, puis nous décrivons en section 4) l'état d'avancement et finissons en section 5) par lister les actions futures.

2. État de l'art

Notre état de l'art s'inscrit principalement dans le domaine de la rétro-ingénierie ou l'ingénierie inverse qui a pour objectif de déterminer le fonctionnement interne d'un système informatique ou son architecture logicielle.

Il existe plusieurs techniques de rétro-conception tel que : « Software Clustering », « Design recovery », « code browsing », « clone detection », etc... [1] Pour savoir quelle technique utiliser il faut définir d'une manière précise ce dont nous disposons comme données d'entrée et ce que nous devons obtenir comme résultats. Dans notre cas nous ne disposons que du code source écrit en langage C# de la technologie Windows .net, et nous avons besoin de modifier et restructurer ce code pour pouvoir intégrer de nouvelles exigences.

La technique de Software Clustering permet de restructurer l'architecture logicielle en se basant uniquement sur le code source. En effet, cette technique consiste à regrouper des entités du code source, principalement des classes, dans des packages ou modules.

L'application du Software Clustering varie selon les spécificités du système étudié, nous pouvons ainsi citer l'algorithme proposé en [2] qui permet de construire un graphe de dépendances entre les classes du système étudié. Dans cet algorithme une classe est dite entité, nous cherchons les dépendances structurelles entre les entités du système pour construire les ensembles suivants :

Soit $C(S)$ l'ensemble des classes de notre système informatique, nous recherchons les quatre relations suivantes :

$$\begin{aligned}(c_1, c_2) \in ECIG &\Leftrightarrow c_1 \text{ extends } c_2 \\(c_1, c_2) \in ECAG &\Leftrightarrow c_1 \text{ aggregates } c_2 \\(c_1, c_2) \in ECUG &\Leftrightarrow c_1 \text{ uses } c_2 \\ESCDG &:= ECIG \cup ECAG \cup ECUG\end{aligned}$$

Le graphe $G_x := (C(S), E_x)$ est appelé graphe de dépendance structurelle de classes (SCDG). Ce graphe nous permet d'identifier les classes ayant des relations fortes entre elles, et de les regrouper, par exemple dans des packages.

3. Problématique

Notre logiciel est composé d'un module de dessin intégré (DAO) et de plusieurs modules métier (Topographie, Terrassement, Route, Réseaux, etc...). Existant depuis une dizaine d'années, ce logiciel n'est pas conçu pour intégrer facilement de nouveaux modules. D'autre part, il permet difficilement de migrer vers une autre plateforme de dessin. C'est pourquoi nous voulons faire évoluer l'application actuelle vers une nouvelle application répondant à des nouvelles exigences : maintenabilité, évolutivité, modularité, interopérabilité avec d'autres moteurs graphique. De ce fait nous posons les questions suivantes :

- Comment retrouver l'architecture du logiciel actuel en absence de documentation ?
- Comment abstraire les modules métier du logiciel actuel sous forme de modèles UML ?

- Quelle architecture logicielle proposer pour le nouveau produit ? Cette architecture doit : permettre l'interopérabilité avec les plateformes de dessin existantes ; faciliter la maintenance ; permettre le découpage en modules indépendants ; permettre de construire facilement des produits 'sur mesure' pour l'utilisateur (l'utilisateur choisit les modules dont il a besoin).
- Quels sont les éléments du logiciel actuel que l'on peut réutiliser dans la reconstruction du système et comment rendre indépendant chaque module métier ?
- Quelle méthodologie employer ?

4. Actions réalisées

Nous avons choisi une approche Ingénierie Dirigée par les Modèles (IDM) [4] pour la rétro-ingénierie d'applications existantes (approche ascendante) et la proposition d'une solution ouverte s'appuyant sur l'ingénierie des modèles (approche descendante) avec pour objectif de modifier l'architecture logicielle et prendre en compte des nouvelles exigences.

Cette approche inspirée de [3], comporte trois phases : La rétro-conception pour remonter du niveau code vers une représentation en modèles UML, et déterminer l'architecture du logiciel actuel ; Ensuite la transformation du code existant, à l'aide ou piloté par ces modèles, avec pour objectif de rendre indépendants les modules métier, prendre en compte la possibilité d'interfaçage avec des plateformes de dessin différentes et assurer la faciliter de maintenance et d'intégration de nouveaux modules métier. Finalement, viendra la phase d'implémentation ou de génération du code en réutilisant au maximum des parties du code existant.

Nous avons commencé la retro-ingénierie par une compréhension fonctionnelle de l'application existante. Ceci nous a permis de définir empiriquement, à partir du découpage fonctionnel, les principaux modules métier. Avec cette vision en tête, nous avons analysé le code afin d'identifier les classes principales et les packages pouvant correspondre à nos modules.

Ensuite, nous avons appliqué l'algorithme proposé en [2] afin de trouver les classes ayant de forts liens structurels. Nous avons alors essayé de regrouper ces classes dans les modules correspondants.

Nous avons dû résoudre le problème posé par les classes ayant des dépendances fortes avec plusieurs classes appartenant à des modules différents : dans quel module les placer ? Nous avons identifié plusieurs stratégies, que nous choisissons après analyse de la classe. Une première solution consiste à diviser la classe originale en plusieurs classes, chacune regroupant les liens structurels avec un module, et à répartir le résultat sur les modules concernés. Une autre solution consiste à mettre la classe originale dans un module 'commun'.

L'étape suivante consiste à identifier les liens entre les modules, et faire en sorte que ceux-ci soient tous unidirectionnels, afin de découpler les modules. Là aussi

nous avons identifié des stratégies afin d'identifier et d'inverser les liens qui sont dans le mauvais sens (interfaces, inversion de dépendances, fabriques ...).

L'utilisation du langage C# .net dans le logiciel existant nous a posé des problèmes : manque d'outils de reverse pour analyser le code; structuration des fichiers ne correspondant pas à la structuration en classes et package. En effet, .net autorise plusieurs classes par fichier, ainsi qu'une classe sur plusieurs fichiers ! De plus, C# n'encourage pas l'utilisation de la notion de « paquetages », rendant l'analyse du code difficile.

5. Actions futures

Pour la suite de notre travail nous avons défini les tâches suivantes :

- Définir l'architecture du futur logiciel répondant à nos nouvelles exigences: interopérabilité avec les moteurs graphique, facilité de maintenance, logiciel 'à la carte'. Pour ces exigences, nous envisageons d'utiliser la notion de « ligne de produits ».
- Continuer la migration de l'application existante vers la nouvelle architecture. Cela implique certainement de découpler plus encore les modules identifiés.
- Fournir une solution permettant l'assemblage des modules tout en gardant leurs indépendances les uns des autres.
- Générer ou implémenter les briques logiciels manquantes à la nouvelle architecture (module d'assemblage, ligne de produit ...).
- Développer des outils d'aide ou une plateforme en vue de l'automatisation de la méthodologie utilisée.

Bibliographies

- [1] Tonello P., Torchiano M., Du Bois B., Systà T., “Empirical studies in RE state of the art and future trends”, *Empir Software Eng* (2007) 12:551–571.
- [2] Fabien B., Stephan D., “On the impact of software evolution on software clustering”, *Empir Software Eng* (2013) 18:970–1004.
- [3] Serge D., Stéphane D., Oscar N, and Square Bracket Associates. Object-oriented Reengineering Patterns. [Bern (Neubrückstrasse 10)]: Square Bracket Associates [c/o O. Nierstrasz], 2008, p. 9-12.
- [4] Jean-Marc J., Benoît C., Didier V., *Ingénierie Dirigée Par Les Modèles Des Concepts à La Pratique*. Paris: Ellipses, 2012.

Vers un enrichissement sémantique des modèles de processus configurables

Loubna El Faquih, Hanae Sbaï

Equipe Qualité des Architectures d'Entreprise, Développement et Intégration (AIQualsadi)

ENSIAS, Université Mohammed V-Souissi, Rabat (10000), Maroc

loubna.elfaquih@um5s.net.ma, sbai.hanae@um5s.net.ma

MOTS-CLÉS : Modèles de processus métier sémantiques, Modèles de processus configurables, Variabilité, Ontologie.

KEYWORDS: Semantic business process model, Configurable process model, Variability, Ontology.

ENCADREMENT : Mounia Fredj (PR)

1. Contexte

La gestion des processus métier (Business Process Management : BPM) consiste à modéliser les processus métier de l'entreprise. L'adoption d'une approche BPM favorise l'alignement des systèmes d'information sur les processus métier. Cependant, les processus ont souvent été utilisés pour documenter le métier des entreprises. Aujourd'hui, le concept de processus métier sémantiques (Hepp *et al.* 05) a été introduit pour enrichir les processus par la sémantique, afin de faciliter la réutilisation, la configuration, et l'adaptation de ces processus. Actuellement, la réutilisation dans l'approche BPM est introduite par le concept de modèle de processus de référence (Thomas *et al.* 06). Cependant, pour automatiser la personnalisation de ces modèles, la solution de *modèle de processus configurable* (MPC) (Rosemann *et al.* 07) a été proposée, pour gérer la variabilité des processus métier. Les travaux existants sur les modèles de processus configurables s'intéressent à la modélisation de la variabilité capturée par le processus, à la configuration et l'évolution de ces processus. Les MPC sont des modèles de processus qui intègrent le concept de la variabilité en regroupant plusieurs variantes du processus dans un seul modèle. L'enrichissement des MPC par la sémantique, et particulièrement l'enrichissement des concepts de la variabilité par la sémantique, peut permettre de concevoir un modèle de processus configurable intelligent. Ceci facilitera sa recherche, sa configuration et aussi son évolution. Ces conclusions sont appuyées par les études de (Valença *et al.* 13) et (La Rosa *et al.* 13) qui ont démontré la nécessité d'automatiser les variantes des processus métier et de fournir un support de prise de décision pour les analystes lors de la configuration et l'évolution des MPC. Notre recherche s'intéresse à l'enrichissement sémantique de ces MPC. Dans la section suivante, nous présentons un état de l'art des travaux connexes à notre recherche.

2. État de l'art

Dans cette section, nous présentons l'état de l'art des travaux sur la gestion des processus métier sémantiques, en plus des principales contributions dans la gestion de la variabilité et la modélisation sémantique du MPC.

- *Les approches de modélisation des processus métier sémantiques*

Pour modéliser les processus métier sémantiques, plusieurs frameworks et méthodes ont été proposés (Hepp *et al.* 05, Weber *et al.* 09). La sémantique a été également introduite au niveau des autres phases du cycle de vie du BPM (configuration, exécution et analyse). L'objectif étant de : i) réduire l'écart entre le métier et la technologie, ii) permettre une meilleure traduction entre les besoins de l'entreprise capturés par les modèles de processus et les besoins réels des parties prenantes.

- *Modèles de processus configurables : Gestion de la variabilité et modélisation sémantique*

Les travaux existants dans le domaine de la gestion de la variabilité des processus métier concernent trois phases principales : la conception, la configuration et l'évolution des MPC. Concernant la conception et la modélisation des MPC, les principales contributions sont axées sur les langages de modélisation de la variabilité : CVL (Ayora *et al.* 12). En outre, d'autres travaux ont discuté l'exactitude (correctness) des variantes relatives à un modèle de processus de référence (Aalst *et al.* 10). La configuration des MPC a été traitée dans différentes études (Hallerback *et al.* 10), la contribution la plus récente de (Huang *et al.* 13) propose d'assister l'utilisateur dans la phase de configuration du MPC en implémentant une ontologie. Finalement, l'évolution des MPC a été étudiée dans (Ayora *et al.* 13) et (Sbai *et al.* 14). Dans ce dernier article, les auteurs ont proposé une méthodologie basée sur les patrons pour guider les concepteurs dans la modélisation et l'évolution des MPC. Nous remarquons que la sémantique est peu traitée par les travaux existants sur les MPC. Après ce bref aperçu de l'état de l'art des axes liés à notre recherche, nous expliquons dans la section suivante la problématique de notre thèse.

3. Problématique

D'après l'état de l'art, la représentation de la sémantique n'a pas été traitée au niveau de la modélisation des MPC, les travaux se limitant à la proposition de quelques contraintes sémantiques pour valider la configuration des MPC. Concernant l'évolution des MPC, nous notons qu'il existe un manque de processus d'évolution intelligent pour gérer le changement des MPC au cours de la conception et de l'exécution. Dans cet article, nous proposons un enrichissement sémantique des modèles de processus configurables, basé sur « une ontologie MPC », comme une première étape vers un processus d'évolution intelligent pour gérer l'évolution des MPC. Un résumé de notre approche est détaillé dans la section suivante.

4. Actions réalisées

Afin d'enrichir sémantiquement les MPC, nous proposons un framework basé sur les ontologies. Une ontologie est définie comme "une spécification formelle explicite d'une conceptualisation partagée d'un domaine d'intérêt" (Gruber *et al.* 93). Le framework contient trois composants principaux: le modèle configurable, l'ontologie MPC (qui formalise les contraintes sémantiques liées à la variabilité du MPC), et l'ontologie de domaine. Afin de lier l'ontologie MPC à l'ontologie de domaine, des règles sémantiques sont utilisées pour traduire les contraintes de l'ontologie de domaine vers l'ontologie. Les composants du framework sont illustrés dans la figure 1.

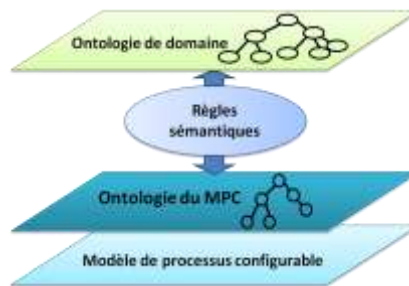


Figure 1: Framework d'enrichissement sémantique des MPC montrant la hiérarchie des classes de l'ontologie

L'ontologie MPC est créée en utilisant le langage OWL2.



Figure 2: Graphe de l'ontologie du MPC

Les différentes règles sémantiques qui capturent la variabilité des MPC sont exprimées par l'ontologie MPC en utilisant les relations entre les classes « object properties » ainsi que les cardinalités. Nous donnons un exemple de règle sémantique : « Une variante dépend d'un seul point de variation » est équivalente au niveau de l'ontologie MPC à « Variant **dependsOn** *exactly 1* Variation_Point ».

5. Actions futures

Dans ce travail, nous avons proposé une ontologie qui capture les concepts sémantiques de la variabilité d'un modèle de processus configurable donné. Nous avons effectué la représentation des concepts ainsi que la détermination des relations. Comme futur travail, nous allons: i) Développer l'ontologie MPC (définition des axiomes et peuplement de l'ontologie). ii) Raffiner les règles

sémantiques en choisissant une ontologie de domaine spécifique. iii) Implémenter le framework. La deuxième étape de notre contribution traitera de l'optimisation du processus d'évolution des MPC en introduisant une approche d'évolution des MPC basée sur l'ontologie du MPC créée.

Bibliographie

- A. Hallerback, T. Bauer, and M. Reichert, "Capturing variability in business process models: the Provop approach", *Journal of Software Maintenance and Evolution: Research and Practice*, 22.6-7 (2010): 519-546.
- C. Ayora, V. Torres, and V. Pelechano, "Variability management in business process models". Technical Report 17, PROS-UPV, 2012.
- C. Ayora, V. Torres, B. Weber, M. Reichert, V. Pelechano, "Enhancing Modeling and Change Patterns". *BMMDS/EMMSAD*, pp: 246-260, 2013.
- G. Valença, C. Alves, V. Alves and N. Niu, "A Systematic Mapping Study On Business Process Variability", *International Journal of Computer Science & Information Technology* 5.1, 2013.
- H. Sbai, M. Fredj and L. Kjiri, "A pattern based methodology for evolution management in business process reuse", *IJCSI International Journal of Computer Science Issues*, Vol. 11, Issue 1, No 1, January 2014.
- I. Weber, J. Hoffmann, J. Mendling and J. Nitzsche, "Towards a Methodology for Semantic Business Process Modeling and Configuration", *Service-Oriented Computing - ICSOC 2007 Workshops*. Springer Berlin Heidelberg, 2009.
- M. Hepp, F. Leymann, C. Bussler, J. Domingue, A. Wahler and D. Fensel, "Semantic Business Process Management: Using Semantic Web Services for Business Process Management", *ICEBE - IEEE Conference on e-Business Engineering* (pp. 535-540), China, 2005.
- M. La Rosa, W.M.P. van der Aalst, M. Dumas, and F.P. Milani, "Business process variability modeling: A survey", *International conference on Business Process Management*, 2013.
- M. Rosemann and W. M. P. van der Aalst, "A configurable reference modelling language". *Information Systems*, 32, 2007.
- O. Thomas, "Understanding the Term Reference Model in Information Systems Research: History, Literature Analysis and Explanation". In *Business Process Management Workshops*, pp. 484-496. Springer Berlin Heidelberg, 2006.
- T. Gruber, « A Translation Approach to Portable Ontology Specifications», *Knowledge Acquisition*, Vol. 5, No. 2, pp. 199-220, 1993.
- W.M.P. van der Aalst, M. Dumas, F. Gottschalk, A.H. ter Hofstede, M. La Rosa and J. Mendling, "Preserving correctness during business process model configuration". *Formal Aspects of Computing* 22.3-4: 459-482, 2010.
- Y. Huang, Z. Feng, K. He, and Y. Huang, "Ontology-Based Configuration for Service-Based Business Process Model". In *Services Computing (SCC), 2013 IEEE International Conference on*, pp. 296-303. IEEE, 2013.