Construction d'une ontologie dans le domaine financier pour la détection de fraudes

Benjamin Auger [1], **Hamza Chergui** [1,2], **Yara Chehade** [1], **Jana El Kadri** [1], **Lylia Abrouk** [1], **Nicolas Cabioch** [2]

1. Laboratoire d'Informatique de Bourgogne - EA 7534
Univ. Bourgogne Franche-Comté
9, Avenue Alain Savary, F-21078 Dijon - France
hamza.chergui,yara_chehade,jana_el-kadry@etu.u-bourgogne.fr
lylia.abrouk,benjamin.auger@u-bourgogne.fr

2. SKAIZen Group
14, rue de Mantes, 92700 Colombes
hchergui,ncabioch@skaizengroup.fr

RÉSUMÉ. *L'objectif de notre travail est de peupler automatiquement à partir de sources hétérogènes une base de connaissances du domaine financier et structurer les informations extraites afin de permettre aux clients d'optimiser leurs moteurs de détection de fraude et de contrôles de sanction. Nous présentons dans cet article notre projet en cours de construction d'une ontologie financière basée sur la base de connaissances KYC et des transactions internationales interbancaires SWIFT aux caractéristiques spécifiques. Notre travail s'inscrit dans le cadre d'un projet de collaboration avec une entreprise dans le domaine financier.*

ABSTRACT. *The goal of our work is to automatically enrich a financial knowledge base from heterogeneous sources and structure the information extracted in order to allow clients to optimize their fraud detection and sanction control engines. We present in this article our project of financial ontology based on the KYC knowledge base and SWIFT international interbank transactions with specific characteristics. Our work is part of a collaborative project with a company in the financial domain.*

MOTS-CLÉS : *Ontologie, détection de fraude, finance, KYC*

KEYWORDS: *Ontology, fraud detection, finance, KYC*

1. Introduction

Dans le cadre des contrôles anti-fraude, la lutte contre le blanchiment d'argent ainsi que le financement du terrorisme, chaque institution financière a pour obligation

d'avoir des systèmes en place pour lutter contre ces activités illégales. Ces systèmes doivent pouvoir analyser des transactions financières pouvant avoir une dimension internationale et interbancaire. C'est dans ce contexte que l'entreprise SKAIZen Group développe un projet de recherche et d'innovation qui a pour but de construire une vision 360° des clients et des contreparties financières visant d'une part à peupler, à partir de sources hétérogènes, une base de faits vérifiés et d'autre part à permettre à leurs clients d'optimiser leurs moteurs de détection de fraudes et de contrôle de sanction associés aux transactions financières. La création d'une ontologie spécialisée sur les informations financières et liées à la conformité bancaire est une première étape dans ce projet de construction d'une base de connaissances. En organisant les données sous forme d'une ontologie spécifiée, nous pourrons interroger la base de connaissances sur les relations entre différentes entités (personnes ou organisations). Notre démarche de construction de base de connaissances consistera en un processus de peuplement d'une ontologie spécialisée dans les contrôles de flux financiers (conformité bancaire, KYC *Know Your Customer*, etc.).

Des ontologies spécialisées existent déjà dans divers domaines scientifiques et sont utilisées pour la description de l'information dans différentes spécialités (formation, biologie, médecine...). (Noy *et al.*, 2008) décrivent un portail d'ontologies biomédicales où (Psyché *et al.*, 2003) discute l'apport des ontologies dans les environnements de formation à distance. Dans la littérature, nous avons également identifié plusieurs ontologies orientées pour l'économie ou la finance. L'ontologie REA (Resource, Evènement, Agent) a été proposée initialement par McCarthy (Mccarthy, 1982) pour les applications de comptabilité. Dans le modèle REA, les Agents comprennent les entités comme organisations ou individus impliquées dans un évènement qui décrit un type de relation économique. Cette ontologie est étendue *REA Business Management Ontology* (Schwaiger, 2016) pour l'entreprise et la gestion d'information.

Dans le domaine financier, la détection de fraudes est basée sur divers informations (transaction, profil client, ...). Ces dernières sont nombreuses, d'ordre qualitatif ou quantitatif et sont de sources hétérogènes, chez plusieurs acteurs du domaines (institutions financières, entreprises, ...). A notre connaissance, il n'existe pas d'ontologie décrivant des transactions bancaires dans le domaine financier et plus particulièrement des transactions interbancaires internationales aux caractéristiques particulières SWIFT. Par conséquent, la construction d'une telle ontologie nous permettrait de décrire ce modèle SWIFT qui vise à devenir une norme ISO 20022 [1] et nous permettre de détecter des fraudes dans ces transactions grâce à la définition de règles métiers explicitant le statut des transactions en fonction des informations dans l'ontologie. Nous présentons dans la section suivante notre travail en cours de construction d'ontologie financière.

1. https://www.iso20022.org/iso-20022-message-definitions

2. L'ontologie financière CTO

L'objectif de notre projet est de construire une ontologie transactionnelle adaptée aux transactions SWIFT. La réalisation d'un tel projet aura un double apport pour notre entreprise car elle pourra se coupler notamment à la fois avec nos travaux sur l'extraction de connaissances d'articles financiers (Jabbari *et al*., 2019 ; 2020) et de nos travaux sur la détection de fraude financière (Chergui *et al*., 2022). En effet, à notre connaissance il n'existe pas de travaux couplant des approches d'apprentissage automatique et ontologique sur la détection de fraude financière. Nous avons pour ambition de proposer à la fois une démarche sur la construction d'une ontologie dans le domaine financier (SWIFT) et une approche sur la démarche d'implémenter une solution avec des techniques d'apprentissage automatique.

Les transactions SWIFT sont des messages répondant à une norme appelée ISO, cette dernière décrit les standards de formalisation de ces messages. Actuellement, ces messages suivent la norme ISO 150022 MT [2], qui sera modifiée dans les prochaines années. En effet, cette norme migrera pour une nouvelle appelée ISO 20022 MX à partir de novembre 2022 où les messages auront la spécificité d'être formalisés à l'aide du langage XML. En outre, les messages MT correspondant à l'ancien format pourront encore être utilisés jusqu'en 2025.
Plusieurs types de messages MT ou MX sont utilisés pour des fonctions précises au sein du réseau SWIFT par lequel circule tous les messages. Pour nos travaux, nous avons décidé de nous intéresser aux virements d'argent internationaux représentés par le type "103" pour les messages MT (norme ISO 150022) et "pacs.008" pour les messages MX (norme ISO 20022). De nombreuses informations sur une transaction sont présentes dans ces messages. Nous avons sélectionné dans un premier temps les informations pertinentes dans le cadre de détection de fraude.

Nous avons choisi de retenir trois classes : la classe **Personne** correspond aux clients présents dans la base de connaissances KYC. La classe **Compte** correspond au compte d'un client bancaire des **personnes**, cette classe permet de faire la jonction entre les personnes présentes dans la base de connaissances et les personnes réalisant des transactions. La classe **Transaction** comporte les attributs montant, devise et date, une transaction est réalisée entre deux **comptes**. Le Tableau 1 montre les relations entre les concepts.

Nous définissons les contraintes de cardinalités suivantes :

– Une personne peut avoir plusieurs comptes bancaires Personne $\subseteq \exists$ possède.Compte

– Un compte n'appartient qu'à 1 seul personne Compte $\subseteq \doteq 1$ inv_possède.Possède

– Un compte effectue 1 à plusieurs transactions (transfert d'argent) Compte $\subseteq \exists$ effectueTransaction.Transaction

2. https://help.flywire.com/hc/fr/articles/360012919014-Qu-est-ce-qu-un-message-SWIFT-MT103-

TABLEAU 1. *Rôles avec leurs domaines*

Rôle	Domain	Range	Rôle inverse
possède	Personne	Compte	inv_possède
effectueTransaction	Compte	Transaction	inv_effectueTransaction
recoitTransaction	Compte	Transaction	inv_recoitTransaction

– Un compte reçoit 1 à plusieurs transactions (reçu d'un transfert d'argent) Compte $\subseteq \exists$ recoitTransaction.Transaction

Nous avons conçu une première version de notre ontologie sous l'outil protégé [3] avec les concepts présents sur la figure 1.

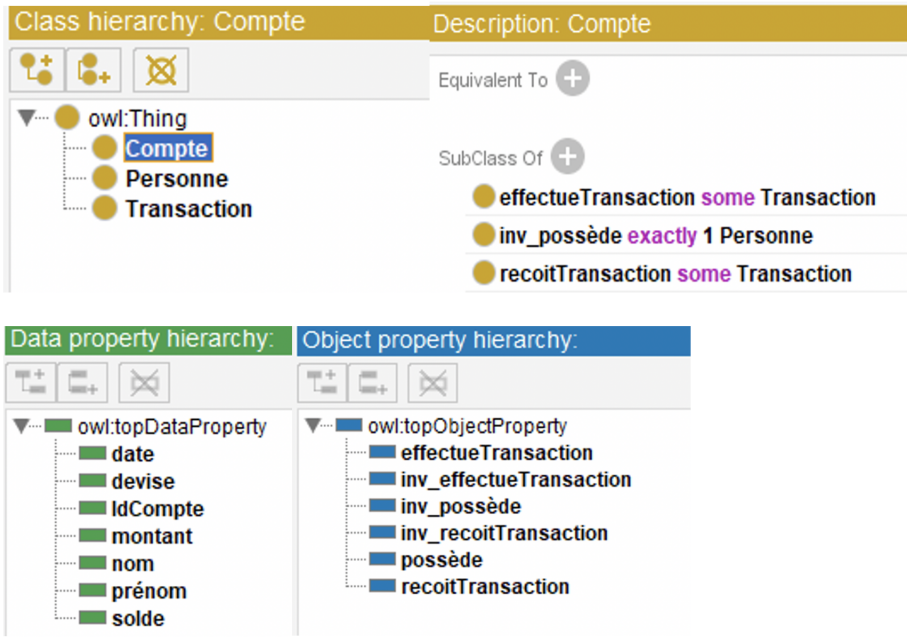

FIGURE 1. *L'ontologie avec l'outil protégé*

Une fois l'ontologie peuplée avec des messages MT et MX, nous avons rédigé quelques règles SWRL (Semantic Web Rule Language) pour la détection de fraudes. Par exemple la règle affichée ci-dessous permet d'obtenir les transactions réalisées

3. https://protege.stanford.edu/

de plus de 13000 dollars par un compte Iranien possédant un solde inférieur à 10000 dollars. Enfin, nous pourrons faire des recherches supplémentaires sur ces transactions à l'aide de notre base de connaissance.

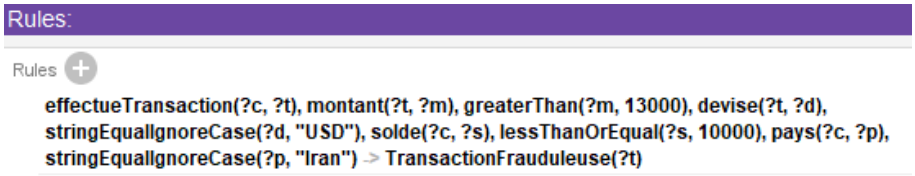

FIGURE 2. *Exemple de règle de détection de fraude*

L'efficacité de la proposition sera évaluée à l'aide d'un jeu de données réelles labellisées obtenu grâce à une collaboration avec l'entreprise SKAIZen Group. Notre ontologie sera enrichie à l'aide de ce jeu de données qui est composé d'un ensemble de transactions. Ces dernières possèdent un label qui nous permet de savoir s'il s'agit d'une transaction frauduleuse ou légitime. Une fois l'ontologie enrichie, nous pourrons faire tourner nos règles de détection de fraude qui sortiront les transactions jugées frauduleuses. Nous pourrons ensuite vérifier à l'aide des labels si les transactions frauduleuses le sont réellement. D'une manière plus globale, nous évaluerons notre proposition à partir des mesures de précision, de rappel et de F1-Score qui nous permettront de mesurer l'efficacité de la proposition.

3. Conclusion et perspectives

Dans le cadre de notre projet en cours, nous mettons en place un cadre formel et un outil permettant la détection de crimes financiers en modélisant le comportement des clients en utilisant les données transactionnelles. La construction de l'ontologie financière CTO est la première étape de ce travail. Dans la suite de ce projet, nous allons développer des règles métier complexes et adaptables en fonction de l'évolution des règles de fraudes et avec l'aide d'experts du domaine. Ce travail sera couplé par la suite avec notre base de connaissances KYC en enrichissant l'ontologie CTO avec le profil des clients à partir de sources de données hétérogènes.

Remerciements Ce travail est soutenu à la fois par l'entreprise SKAIZen Group, l'ANRT et l'ANR (France Relance).

Bibliographie

Chergui H., Alfred R. A., Abrouk L., Jabbari A., Cullot N. (2022). Détection d'anomalies: une méthode appliquée aux transactions interbancaires. *Extraction et Gestion des Connaissances: EGC'2022*, vol. 38.

Jabbari A., Sauvage O., Cabioch N. (2019). Towards a knowledge base of financial relations: Overview and project description. In *2019 ieee second international conference on artificial intelligence and knowledge engineering (aike)*, p. 313–316.

Jabbari A., Sauvage O., Zeine H., Chergui H. (2020). A french corpus and annotation schema for named entity recognition and relation extraction of financial news. In *Proceedings of the 12th language resources and evaluation conference*, p. 2293–2299.

Mccarthy W. (1982, 07). The rea accounting model: A generalized framework for accounting systems shared data environment. *The Accounting Review*.

Noy N. F., Shah N., Dai B., Dorf M., Griffith N., Jonquet C. *et al*. (2008). Bioportal: A web repository for biomedical ontologies and data resources. In C. Bizer, A. Joshi (Eds.), *Proceedings of the poster and demonstration session at the 7th international semantic web conference (iswc2008), karlsruhe, germany, october 28, 2008*, vol. 401. CEUR-WS.org. Consulté sur http://ceur-ws.org/Vol-401/iswc2008pd_submission_25.pdf

Psyché V., Mendes O., Bourdeau J. (2003, 01). Apport de l'ingénierie ontologique aux environnements de formation à distance. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, vol. 10.

Schwaiger W. S. A. (2016). Rea business management ontology: Conceptual modeling of accounting, finance and management control. In *Caise forum.*