



école doctorale **sciences pour l'ingénieur et microtechniques**

THESE DE DOCTORAT EN INFORMATIQUE

Titre : Études et intégration des systèmes multiagents et des ontologies pour la génération de documents contextualisés dans des systèmes complexes

Laboratoire d'accueil : Laboratoire Connaissance et Intelligence Artificielle Distribuées (CIAD) – Université de Technologie de Belfort-Montbéliard - 13 rue Ernest Thierry-Mieg, 90010, Belfort Cedex, France - <http://www.ciad-lab.fr>

Entreprise d'accueil du doctorant pour la thèse (contrat CIFRE) :

ARMTEK - Strasbourg, France – <http://www.armtek.fr>

Encadrement académique de la thèse dans le laboratoire CIAD :

- Directeur de thèse : Prof. Dr. Stéphane GALLAND <stephane.galland@utbm.fr>

- Co-directeur de thèse : Prof. Dr. Christophe NICOLLE <christophe.nicolle@u-bourgogne.fr>

Encadrement technique de la thèse dans la société ARMTEK :

- M. Vincent OSTERMANN <vincent.ostermann@armtek.fr>

I. Sujet de thèse

Mots-clefs : modélisation sémantique de corpus de documents, construction contextuelle et automatique de documents, intelligence artificielle, systèmes multiagents, ingénierie des connaissances

Descriptif détaillé du sujet de thèse :

1. Contexte et problématique principale

L'acquisition des processus par le contenu connaît un engouement très fort depuis quelques années. Cette technologie fait référence à la manière dont les individus acquièrent de nouvelles connaissances et compétences en apprenant le contenu d'un domaine spécifique [1]. Cette approche de l'apprentissage humain se concentre sur la façon dont les apprenants traitent l'information et l'utilisent pour construire de nouveaux processus cognitifs. Selon cette approche, les apprenants acquièrent des processus cognitifs en acquérant le contenu d'un domaine spécifique, tels que des procédures, des règles, des concepts et des principes. Durant ce processus, les apprenants développent également des compétences cognitives telles que la résolution de problèmes, la prise de décisions, la pensée critique et la créativité [2]. Ces compétences sont intégrées dans leur processus de pensée et deviennent disponibles pour être appliquées dans d'autres contextes et situations. Dans ce contexte, la numérisation des processus d'acquisition par le contenu est fondamentale. Le marché mondial des logiciels dédiés a augmenté de 63% en 2018,

ce qui en fait le segment du marché des logiciels d'entreprise qui connaît la plus forte croissance, selon Gartner. De nombreux éditeurs de logiciels, tels que SAP, UiPath, IBM ou Microsoft, fournissent des solutions logicielles. Toutefois, il ne s'agit pas de considérer uniquement la technologie pour répondre aux problèmes posés par la modélisation des connaissances métier et la construction de processus permettant à l'apprenant d'acquérir de nouvelles connaissances. Notamment, un axe de recherche récent consiste à utiliser l'intelligence artificielle, incluant les systèmes multiagents, l'ingénierie des connaissances et l'apprentissage artificiel, afin de construire des solutions innovantes « de méthode par l'acquisition des processus ». Ces nouveaux outils répondront aux problématiques d'automatisation des tâches répétitives qui nécessitaient auparavant une intervention humaine.

Dans le cadre des activités de la société ARMTEK liées à la construction et la maintenance de procédures industrielles dans le milieu de la Défense, il s'agit de concevoir un système permettant de numériser les éléments atomiques (ou atomes) constituant les descriptions des procédures industrielles, et de construire dynamiquement de nouvelles procédures en tenant compte de la sémantique et la forme de ces atomes, ainsi que du profil de la personne devant lire la documentation. Dans ce contexte général, il est proposé de construire un **système d'intelligence artificielle hybride, liant l'ingénierie des connaissances et les systèmes multi-agents**, afin de construire automatiquement et proposer à un utilisateur final un document décrivant une procédure industrielle tout en intégrant la cohérence sémantique des éléments constituant la documentation et le profil utilisateur.

D'une part, le domaine scientifique de l'intelligence artificielle possède un champ d'études concernant l'ingénierie des connaissances dans différents domaines tels que l'industrie [3], les villes intelligentes [4], et également l'énergie [5,6]. Ces différents travaux ont utilisé des techniques issues de la modélisation à base d'ontologies, de réseaux sémantiques, de processus de modélisation, etc. De plus, des travaux se sont intéressés plus récemment à l'évolution au cours du temps des connaissances métiers afin de limiter la part de compétences dans les entreprises [7] ou de s'adapter à de nouvelles informations technico-légales [8] et notamment le domaine du « lifelong machine learning » [9]. Dans le cadre de cette thèse de doctorat, il sera question d'intégrer les derniers résultats de travaux de recherche en termes de modélisation des connaissances métiers afin d'obtenir un modèle sémantiquement enrichi pouvant être utilisé pour la construction de nouveaux processus par un ensemble d'agents autonomes.

D'autre part, le domaine de la modélisation de systèmes complexes à base d'intelligence artificielle (comme les systèmes multi-agents), c'est intéressé à la modélisation des concepts, processus et méthodes pour construire des méthodologies de création de systèmes complexes. Or, les systèmes étudiés par ARMTEK sont par nature complexes. Il serait par conséquent intéressant d'explorer cette direction de recherche afin de formaliser la méthodologie utilisée par les collaborateurs de ARMTEK et ses clients et de permettre sa généralisation et son évolution en fonction des connaissances et des pratiques technico-légales. Dans ce cadre, il est important d'étudier les modèles existants. Ainsi, le problème de création automatique d'un document technique n'est pas récent et n'est pas spécifique au domaine de l'industrie. Par exemple, Andersen et al. [10] discutent de la documentation technique des grands systèmes d'équipements. Dans le cadre de cette thèse, il s'agit de construire une méthodologie innovante complétant celles déjà existantes et généralement basées sur les étapes suivantes [11, 12, 13] :

- Modélisation des connaissances métier, généralement en utilisant des outils sémantiques et d'ontologie ;
- Calcul de plusieurs bonnes solutions qui répondent aux exigences des clients ;
- Sélection des meilleures options parmi l'ensemble des bonnes solutions ;
- Construction du document.

Une minorité des articles issus de la littérature utilisent l'ingénierie d'ontologie ou des solutions de calcul basées sur l'IA et plus particulièrement sur l'intelligence artificielle distribuée ou les systèmes multiagents afin de répondre aux défis de la génération automatique, contextuelle et intelligente des documents, constituant ainsi une direction de recherche encore ouverte. Les systèmes multiagents et l'ingénierie des connaissances offrent plusieurs avantages pour la création et gestion autonome et dynamique de processus, notamment une intelligence distribuée, une interopérabilité, une gestion de la complexité, une adaptabilité, une sémantique commune, une réutilisation des connaissances, un raisonnement, une flexibilité et une facilitation de l'apprentissage de l'utilisateur.

Malgré un certain nombre de travaux scientifiques en relation avec le contexte et les problématiques de cette thèse (analyse par modèle de survie de l'introduction d'un système de gestion intégré du travail, la modélisation et gestion de flux par systèmes multi-agents, ...), des verrous scientifiques persistent :

- 1. Comment modéliser un processus métier spécifique incluant de multiples acteurs ?**
- 2. Comment modéliser le comportement de chaque élément participant à l'évolution du système complexe ?**
- 3. Comment interfacier les opérateurs existants sur le terrain avec les modules de gestion des agents dans les systèmes multiagents pour l'automatisation du processus ?**
- 4. Comment améliorer l'acceptation du changement par les acteurs locaux lors de l'introduction d'un nouveau système de gestion ?**

Aujourd'hui, la génération dynamique de documents est largement répandue avec le développement des IA génératives à base de transformers comme les approches LLM (Large Language Model) dont le principal outil est ChatGPT. Néanmoins, de nombreux travaux pointent les limites de ces technologies et des résultats attendues. L'utilisation d'approches multiagents combinées à des ontologies pour la génération de documents, en particulier dans le contexte de l'adaptation contextuelle des processus métiers, présente plusieurs avantages par rapport aux approches basées sur des LLM comme ChatGPT. Ainsi, les limitations principales des LLM qui justifient l'adoption d'une approche multiagents sont : (i) Les LLM, bien qu'excellents pour générer du contenu basé sur des patrons linguistiques, ont des limites en termes de compréhension contextuelle profonde et dynamique, par conséquent ils peuvent manquer de la précision spécifique au domaine nécessaire pour certains processus métiers. En particulier, les LLM intègrent et appliquent difficilement des règles et politiques métier complexes de manière cohérente ce qui peut entraîner des réponses erronées ou peu pertinentes en présence d'informations incomplètes ou ambiguës ; (ii) À cela s'ajoute des contraintes de mise à jour des LLM, ces derniers ayant, par leur approche d'apprentissage des limites en termes d'interactivité et d'adaptabilité en temps réel.

Pour répondre à ces limites, nous avons besoin de développer une nouvelle approche de raisonnement pour mieux comprendre et réagir aux changements dans les processus métiers et les exigences contextuelles spécifiques. Cette approche permettrait une adaptation plus précise et pertinente, en particulier sur la représentation et gestion des règles et contraintes métier, assurant ainsi que les documents générés respectent les normes et réglementations spécifiques tout en s'adaptant en temps réel aux changements de l'environnement ou aux demandes des utilisateurs. Notre hypothèse scientifique au regard de notre expertise du domaine de l'intelligence artificielle et de son développement scientifique, serait de concevoir une approche basée sur des systèmes multiagents, en interaction avec des ontologies. Par leur capacité d'évoluer par simulation, ainsi que de modélisation et raisonnement sur des vérités d'experts (ontologie), cette approche pourrait répondre aux limites présentées ci-dessus. Les systèmes

multiagents, grâce à leur capacité de raisonnement distribué et notamment à prendre des décisions basées sur des règles logiques et des connaissances structurées (ontologie), offriront une meilleure adaptation contextuelle, une personnalisation spécifique au domaine, et une gestion plus précise des contraintes et des exigences complexes des processus métiers, ce qui est essentiel pour la génération de documents dans des systèmes complexes.

2. Objectif général de la thèse

L'objectif de cette thèse est de proposer un modèle et les outils associés afin de répondre à une ou plusieurs des problématiques scientifiques citées ci-dessus. Une approche de modélisation du système se rapprochant de techniques d'optimisation par méta-heuristique, telles que la Particle Swarm Optimisation [14] ou des modèles inspirés des colonies d'insectes [15, 16, 17] est envisagée. Toutefois, il convient de comprendre que l'espace de recherche des solutions étant constitué par un graphe sémantique, les approches existantes doivent être adaptées. De plus, l'heuristique d'exploration du graphe qui est appliquée par les agents doit être spécifiée afin d'y intégrer les contraintes métier et les informations d'usage liées au domaine d'application spécifique d'ARMTEK.

L'utilisateur final sera représenté par un agent dans le système afin de permettre la mise en œuvre de méthodes de coopération, coordination, négociation entre cet agent et les différents agents représentant les documents possibles. Ceci permettra au système de construire plusieurs solutions et de sélectionner celle répondant au mieux aux attentes de l'utilisateur final. Le caractère innovant du projet est de concevoir ce système multiagent hiérarchique où la stratégie de collaboration des agents réside dans l'analyse de la complémentarité sémantique de l'information qu'ils portent et sa distance de pertinence avec le profil et le contexte du lecteur.

Références bibliographiques :

[1] Club des Pilotes de processus. "Guide de la gestion des processus : BPM CBOK V3 amélioré au niveau européen et traduit par les membres du Club des Pilotes de processus Ed. 2." (2019).

[2] Debarnot, Ursula. "L'influence du sommeil sur le processus de consolidation des mémoires procédurales cognitive et perceptivo-motrice." (2008). Thèse.

[3] Davy MONTICOLO, Simona MIHAITA, Hind DARWICH, Vincent HILAIRE. "An agent-based system to build project memories during engineering projects." In Knowledge-Based Systems, vol 68, pp. 88-102, dec 2014. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2013.12.022.

[4] Philippe DESCAMPS, Vincent HILAIRE, Olivier LAMOTTE, Sebastian RODRIGUEZ. "An experience of Engineering of MAS for Smart Environments: extension of ASPECS." In Smart Digital Future 2016, jun 2016.

[5] Jiawei ZHU, Fabrice LAURI, Abderrafiaa KOUKAM, Vincent HILAIRE, Yishuai LIN, Youquan LIU. "A hybrid intelligent control based cyber-physical system for thermal comfort in smart homes." In International Journal of Ad Hoc and Ubiquitous Computing, vol 30, pp. 199-214, dec 2019.

[6] Gillian BASSO, Massimo COSENTINO, Vincent HILAIRE, Fabrice LAURI, Sebastian RODRIGUEZ, Valeria SEIDITA. "Engineering multi-agent systems using feedback loops and holarchies." In Engineering Applications of Artificial Intelligence, vol 55, pp. 14-25, jun 2016.

[7] Candanedo, I. S., Nieves, E. H., González, S. R., Martín, M. T. S., & Briones, A. G. (2018, August). Machine learning predictive model for industry 4.0. In International Conference on Knowledge Management in Organizations (pp. 501-510). Springer, Cham.

[8] Dalal, S. R., Shekelle, P. G., Hempel, S., Newberry, S. J., Motala, A., & Shetty, K. D. (2013). A pilot study using machine learning and domain knowledge to facilitate comparative effectiveness review updating. Medical Decision Making, 33(3), 343-355.

- [9] Chen, Z., & Liu, B. (2018). Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 12(3), 1-207.
- [10] Andersen, A., et Munch, K. H. (1991). Automatic generation of technical documentation. *Expert Systems with Applications*, 3(2):219–227.
- [11] Chen, Y., Hong, T., et Piette, M. A. (2017). Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Applied Energy*, 205:323–335.
- [12] Mrasek, R., Mülle, J., et Böhm, K. (2015). Automatic generation of optimized process models from declarative specifications. In Zdravkovic, J., Kirikova, M., et Johannesson, P., editors, *Advanced Information Systems Engineering*, pages 382–397, Cham. Springer International Publishing
- [13] Barba, I., Del Valle, C., Weber, B., et Jimenez, A. (2013). Automatic generation of optimized business process models from constraint-based specifications. *International Journal of Cooperative Information Systems*, 22(02):1350009.
- [14] Russell Eberhart and James Kennedy. Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, volume 4, pages 1942–1948, 1995.
- [15] Marco Dorigo and Thomas Stützle. *Ant colony optimization for np-hard problems*. 2004.
- [16] Dervis Karaboga et al. An idea based on honey bee swarm for numerical optimization. Technical report, Technical report-tr06, Erciyes university, engineering faculty, 2005
- [17] Xin-She Yang. Firefly algorithms for multimodal optimization. In *Stochastic Algorithms: Foundations and Applications: 5th International Symposium, SAGA 2009, Sapporo, Japan, October 26-28, 2009. Proceedings 5*, pages 169–178. Springer, 2009.

II. Modalités de candidatures

Profil demandé pour les candidats en thèse :

- Un diplôme de master ou équivalent en informatique ou autre spécialité en relation avec le sujet de thèse (pour les candidats en dernière année de Master ou de diplôme d'Ingénieur)
- Un très bon niveau en modélisation et programmation informatique (modélisation et programmation orientée objet).
- Des connaissances en intelligence artificielle et plus particulièrement dans les systèmes multiagents ou l'ingénierie des connaissances seront appréciées
- Une bonne maîtrise de l'anglais (oral et écrit) est exigée

Financement et durée de la thèse :

Contrat à durée déterminée de 3 ans dans la société ARMTEK, selon les modalités d'un financement de type CIFRE (<https://www.enseignementsup-recherche.gouv.fr/fr/les-cifre-46510>)

Dates importantes :

- Dossier à envoyer pour le : 29 Février 2024
- Pour les dossiers retenus, entretiens et auditions avec le laboratoire CIAD et l'entreprise ARMTEK : Mars 2024
- Date estimée de début de la thèse : Septembre ou octobre 2024

Candidatures :

Les candidatures doivent être envoyées par email à :

- Prof. dr. Stéphane GALLAND <stephane.galland@utbm.fr>

Le dossier de candidature doit contenir :

- un CV détaillé (**obligatoire**)
- une copie du diplôme de Master ou tout document attestant du niveau de Master (**obligatoire**)
- une copie des bulletins de notes de Master/Cursus Ingénieur (**obligatoire**)
- une lettre de motivation concernant l'inscription en thèse et l'intégration de la thèse dans le projet de carrière (**obligatoire**)
- si possible, une à deux lettres de recommandation de professeurs ou chercheurs

Tout dossier incomplet sera rejeté.