DOCTORAL THESIS POSITION IN COMPUTER SCIENCE

Title: Studies and integration of multiagent systems and ontologies for the generation of contextualized documents in complex systems

Host laboratory: Distributed Knowledge and Artificial Intelligence Laboratory (CIAD) – Belfort-Montbéliard University of Technology - 13 rue Ernest Thierry-Mieg, 90010, Belfort Cedex, France - http://www.ciad-lab.fr

Host company for the doctoral student for the thesis (CIFRE contract):

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

Academic supervision of the thesis in the CIAD laboratory:
- Thesis director: Prof. Dr. Stéphane GALLAND <stephane.galland@utbm.fr>
- Thesis co-supervisor: Prof. Dr. Christophe NICOLLE <christophe.nicolle@u-bourgogne.fr>

Technical supervision of the thesis in the ARMTEK company:
- Mr. Vincent OSTERMANN <vincent.ostermann@armtek.fr>

I. Thesis subject

Keywords: semantic modeling of documents, contextual and automatic document construction, artificial intelligence, multiagent systems, knowledge engineering

Detailed description of the thesis subject:

1. Context and main problem
The acquisition of processes through content has become very popular in recent years. This technology refers to how individuals acquire new knowledge and skills by learning content in a specific domain [1]. This approach to human learning focuses on how learners process information and use it to construct new cognitive processes. According to this approach, learners acquire cognitive processes by acquiring the content of a specific domain, such as procedures, rules, concepts, and principles. During this process, learners also develop cognitive skills such as problem solving, decision-making, critical thinking, and creativity [2]. These skills are integrated into their thought processes and are available for application in other contexts and situations. In this context, the digitalization of acquisition processes through content is fundamental. The global dedicated software market grew 63% in 2018, making it the fastest growing segment of the enterprise software market, according to Gartner. Many software companies, such as SAP, UiPath, IBM, or Microsoft, provide software solutions. However, it is not sufficient to consider technology alone to address the problems posed by modeling business knowledge and building processes that enable

the learner to acquire new knowledge. A recent area of research consists of the use of artificial intelligence, including multi-agent systems, knowledge engineering, and machine learning, to build innovative solutions based on the "process acquisition method". These new tools will address the issues of automating repetitive tasks that previously required human intervention.

As part of the activities of the ARMTEK company linked to the construction and maintenance of industrial procedures in the Defense sector, the aim is to design a system that makes it possible to digitize the atomic elements constituting the descriptions of industrial procedures, and to dynamically construct new procedures taking into account the semantics and the shape of these atoms, as well as the profile of the person who must read the documentation. In this general context, it is proposed to build a **hybrid artificial intelligence system, linking knowledge engineering and multi-agent systems**, to automatically construct and offer to an end-user a document describing an industrial procedure while integrating the semantic coherence of the elements constituting the documentation and the user profile.

On the one hand, the scientific field of artificial intelligence has a subfield related to knowledge engineering and modeling in different fields such as industry [3], smart cities [4], and also energy [5,6]. These different works use techniques based on ontologies, semantic networks, modeling processes, etc. Furthermore, more recently focused on the evolution over time of business knowledge to share skills in companies [7] or adapt to new technical legal information [8] and notably the field of 'lifelong machine learning" [9]. As part of this doctoral thesis, it will be a question related to the integration of the latest research results in terms of knowledge modeling to obtain a semantically enriched model that can be used for the construction of new processes by a set of autonomous agents.

On the other hand, the field of complex systems modeling based on artificial intelligence (such as multi-agent systems) is interested in the modeling of concepts, processes, and methods to build methodologies. The systems considered by ARMTEK are by nature complex. Therefore, it would be interesting to explore this research direction to formalize the methodology used by ARMTEK collaborators and its partners and to allow its generalization and evolution according to knowledge and technical legal practices. In this context, it is important to study existing models. The problem of automatic creation of technical documents is not recent and is not specific to the field of industry. For example, Andersen et al. [10] discuss the technical documentation of large equipment systems. As part of this thesis, the aim is to construct an innovative methodology that complements those already existing and is generally based on the following steps [11, 12, 13]:
- Business knowledge modeling, generally using semantic and ontology tools;
- Calculation of several good solutions that meet customer requirements;
- Selection of the best options from the set of good solutions;
- Construction of the document.

A minority of articles in the literature use ontology or calculation solutions based on AI, and more particularly with multi-agent systems, to respond to the challenges of automatic, contextual, and smart documents. It remains an open research direction. Multi-agent systems and knowledge modeling offer several advantages for the autonomous and dynamic creation and management of processes, including distributed intelligence, interoperability, complexity management, adaptability, common semantics, knowledge reuse, reasoning, flexibility, and facilitation of user learning.

Despite a certain number of scientific works in relation to the context and the issues of this thesis, scientific obstacles persist:
1. **How to model a specific business process that includes multiple actors?**

2. **How can we model the behavior of each element participating in the evolution of the complex system?**
3. **How to interface existing field operators with agent management modules in multi-agent systems for process automation?**
4. **How can improve the acceptance of change by local stakeholders when introducing a new management system?**

Today, the dynamic generation of documents is widespread with the development of generative AI based on transformers such as LLM (Large Language Model) approaches whose main tool is ChatGPT for example. However, numerous works point out the limitations of these technologies and the expected results. The use of multi-agent approaches combined with ontologies for document generation, especially in the context of contextual adaptation of business processes, has several advantages over LLM-based approaches. Thus, the main limitations of LLMs which justify the adoption of a multi-agent approach are: *(i)* LLMs, although excellent for generating content based on linguistic patterns, have limitations in terms of deep and dynamic contextual understanding; therefore, they may lack the domain-specific precision needed for certain business processes. In particular, LLMs have difficulty integrating and applying complex business rules and policies in a coherent manner, which can lead to erroneous or irrelevant answers in the presence of incomplete or ambiguous information; *(ii)* Added to this are the constraints of updating LLMs, the latter having, through their learning approach, limits in terms of interactivity and adaptability in real time.

To address these limitations, we need to develop a new reasoning approach to better understand and respond to changes in business processes and specific contextual requirements. This approach would allow for a more precise and relevant adaptation on the representation and management of business rules and constraints, thus ensuring that the documents generated respect specific standards and regulations while adapting in real time to changes in the environment or to user requests. Our scientific hypothesis, given our expertise in the field of artificial intelligence, would be to design an approach based on multiagent systems with ontologies. By its ability to evolve through simulation, as well as modeling and reasoning on expert truths (ontology), this approach could respond to the limits presented above. Multi-agent systems, thanks to their capacity for distributed reasoning and to make decisions based on logical rules and structured knowledge (ontology), would offer better contextual adaptation, domain-specific personalization, and more precise management of constraints and complex business process requirements, which is essential for document generation in complex systems.

2. General objective of the thesis

The objective of this thesis is to propose a model and associated tools to respond to one or more of the scientific issues cited above. A system modeling approach such as metaheuristic optimization techniques, such as particle swarm optimization [14] or models inspired by insect colonies [15, 16, 17], could be considered. However, because the solution search space is a semantic graph, existing approaches must be adapted. In addition, the graph exploration heuristic which is applied by the agents must be specified and integrate business constraints and usage information linked to the specific application domain of ARMTEK.

The end-user will be represented by an agent in the system to allow the implementation of methods of cooperation, coordination, and negotiation between this agent and the different agents representing the possible documents. This will allow the system to build several solutions and select the one that best meets the end-user's expectations. The innovative nature of the project is to design this hierarchical multi-agent system where the collaboration strategy of the agents lies in the analysis of the semantic complementarity of the information they carry and its distance of relevance with the profile and context of the reader.

Bibliographic References:

[1] Process Pilots Club. 'Guide to process management: BPM CBOK V3 was improved at the European level and translated by members of the Process Pilots Club Ed. 2." (2019).

[2] Debarnot, Ursula. "The influence of sleep on the process of consolidation of cognitive and perceptual-motor procedural memories." (2008). Thesis.

[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, June 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 COSSENTINO, 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, June 2016.

[7] Candanedo, IS, Nieves, EH, González, SR, Martín, MTS, & Briones, AG (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, SR, Shekelle, PG, Hempel, S., Newberry, SJ, Motala, A., & Shetty, KD (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., and Munch, KH (1991). Automatic generation of technical documentation. Expert Systems with Applications, 3(2):219–227.

[11] Chen, Y., Hong, T., and 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., and Böhm, K. (2015). Automatic generation of optimized process models from declarative specifications. In Zdravkovic, J., Kirikova, M., and Johannesson, P., editors, Advanced Information Systems Engineering, pages 382–397, Cham. Springer International Publishing

[13] Barba, I., Del Valle, C., Weber, B., and 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.

Profile requested for thesis candidates:

- A master's degree or equivalent in computer science or other specialty related to the thesis subject (for candidates in the final year of a Master's or Engineering degree).

- A very good level in modeling and computer programming (object-oriented modeling and programming).

- Knowledge in artificial intelligence (not only machine learning) and more particularly in multi-agent systems or knowledge engineering will be appreciated.

- Fluent in English (oral and written) is required.

- Because the application domain of ARMTEK is in the Defense section, the candidate should have EU nationality.

Funding and duration of the thesis:

Fixed-term contract of 3 years in the company ARMTEK, according to the terms of CIFRE type financing (https://www.enseignementsup-recherche.gouv.fr/fr/les-cifre-46510). This contract could be extended at the end of the PHD thesis period, after negotiating with ARMTEK.

Important dates:
- File to be sent by: 10 May 2024
- For the selected files, interviews with the CIAD laboratory and the ARMTEK company: end of May 2024
- Estimated start date of the thesis: September to November 2024

Applications:

Applications must be sent by email to:
- Prof. Dr. Stéphane GALLAND < stephane.galland@utbm.fr >

The application file must contain:
- a detailed CV (**mandatory**)
- a copy of the Master degree or any document attesting to the Master level (mandatory);
- a copy of the Master/Engineering courses that the candidate has obtained during his/her stidues (**mandatory**)
- a cover letter about application to the thesis position and the project of integration and career (**mandatory**)
- If possible, one or two recommendation letters from professors or researchers.

Incomplete applications will automatically be rejected.