PhD proposal in Computer Science
TeleChatBot: a Conversational Agent for Requirements Collect & Analysis to Set up Specification Models
– Application to the design of Telehealth Applications –

Keywords

Requirements Engineering (RE), Artificial Intelligence and Natural Language Processing (NLP), Systems Modelling (UML, SysML).

Description of the thesis project

Context.
It's a well-known fact that software systems development is very often more expensive than expected, development time is difficult to evaluate and the end result disappoints stakeholders. One of the reasons for this is to be found in the upstream system design phase, known as Requirements Engineering (RE), which consists of gathering the expectations of project stakeholders from functional and non-functional points of view. Conventional methods associated to this stage are based on interviews, conducted by the project manager or a RE expert. They take the form of contractual documents in text format, which are difficult to verify and often ambiguous or incomplete. These discussions are therefore fundamental to building the future software system. In addition to interviews, rigorous approaches (goal models, prototyping, definition of user interfaces, use cases, context formalization) lead to set up requirements models as class or requirements diagrams [1], goal model [2], or social models [3]. Unfortunately, software engineering companies often neglect the RE stage, and don't take the necessary care with it.

Issues.
Collecting requirements focuses on understanding the context of the project to be developed and its objectives. It is above all an exercise based on natural language expression and comprehension. It has to be combined to a know-how about the required level of abstraction, so as not to get lost in details that would be more appropriate to the realization stage than the design stage. This know-how is also linked to two areas of expertise: expertise in RE (requirements collect, but also verification and validation) and knowledge about the target domain.
In order to encapsulate the required expertise within a single entity, we wish to study the possibility of automating the collect of requirements and their validation/verification by a conversational agent (chatbot).

Solution to be explored.
Recent advances in AI techniques such as classification and NLP, based on Large Language Models, demonstrate the interest of such techniques for requirements elicitation and analysis [4 - 6]. Some work has been carried out on the improvement of requirements already formulated

using user opinion classification techniques [7-9], and conversational agents are beginning to emerge [10].

As the results are far from satisfactory [11], and the use of AI techniques poses responsibility problems [12], we propose to study how to combine NLP techniques with the capitalization of a know-how relating to the RE and to the application domain (via an ontology, for example). This solution therefore requires experimentation with NLP and modelling techniques, and consequently a good knowledge of both fields.

This work will be implemented and tested as part of a telehealth application design phase.

Objectives and expected results.

Several deliverables are expected from this thesis work:

- an analytical and critical study of the state of the art in both RE and NLP.
- an original method for improving not only the quality of requirements repositories and specification models, but also the process for developing them.
- a software prototype allowing the implementation of the method
- a proof of concept on a telehealth application
- publication of the work developed during the thesis an international conferences or international journals.

References.

[1] "OMG SYSML", version 1.6. *OMG Document*, number formal/19-11-01, Nov. 2019.

[2] A. Van Lamsweerde, E. Letier. "From object orientation to goal orientation: A paradigm shift for requirements engineering," *International Workshop on Radical Innovations of Software and Systems Engineering in the Future*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[3] E.S. Yu. "Social modeling and I*," *Lect. Notes Comput. Sci*, vol. 5600 LNCS, pp. 99–121, 2009.

[4] K. Liu, S. Reddivari, and K. Reddivari. "Artificial Intelligence in Software Requirements Engineering: State-of-the-Art," *Proc. - IEEE 23rd Int. Conf. Inf. Reuse Integr. Data Sci,* pp. 106–111, 2022.

[5] K. Papapanos, J. Pfeifer. "A literature review on the impact of artificial intelligence on in requirements Elicitation and Analysis," *Master report*, University Stockholm, 2023.[2]

[6] S. Das, N. Deb, A. Cortesi, N. Chaki. "Extracting goal models from natural language requirement specifications," in *Journal of Systems and Software*, 2024, vol. 211

[7] J. Wei, A. Courbis, T. Lambolais, B. Xu, P. L. Bernard, and G. Dray. "Zero-shot Bilingual App Reviews Mining with Large Language Models," *35th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2023.

[8] J. Wei, A. L. Courbis, T. Lambolais, P. L. Bernard, and G. Dray. "Towards boosting Requirements Engineering of a Health Monitoring App by Analysing Similar Apps: A Vision Paper," *Proc. IEEE Int. Conf. Requir. Eng.*, pp. 75–80, 2022.

[9] P. Harth, O. Jähde, S. Schneider, N. Horn, and R. Buchkremer. "From Data to Human-Readable Requirements: Advancing Requirements Elicitation through Language-Transformer-Enhanced Opportunity Mining," in *Algorithms*, vol. 16, no. 9, 2023.

[10] Walid Maalej. "From RSSE to BotSE: Potentials and Challenges Revisited after 15 Years," *IEEE/ACM 5th International Workshop on Bots in Software Engineering (BotSE),* Melbourne, Australia, 2023, pp. 19-22.

[11] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Ban, J. Ye, A. Madotto, P. Fung. "Survey of Hallucination in Natural Language Generation," in *ACM Comput. Surv.*, vol. 55, no. 12, 2023.

[12] W. Maalej, Y. D. Pham and L. Chazette, "Tailoring Requirements Engineering for Responsible AI," in *Computer*, vol. 56, no. 4, pp. 18-27, April 2023.

Applicant profile

Required diploma: Master degree in Computer Science.

Technical skills:
- Major Knowledge on Software Development Engineering and System Design
- Knowledge and practice in Model Driven Engineering: Meta Modelling, System Modelling and Model Transformation,
- Basic Knowledge in Machine Learning,
- JAVA or Python programming,

Soft skills:
- Solid oral and written communication capacity in French or in English
- Curiosity and Adaptability capacities to deal with the application domain: understanding knowledge and reasoning of experts in the telehealth domain.

Applicancy procedure

On the the official website of the doctoral school: https://edi2s.umontpellier.fr

Select: Procedures/Recruitements/phD subjects proposals/speciality Informatique

Select "TeleChatBot" in the list + Button Candidate at the bottom.

Deadline: May, 12th at 23h59 (CEST)

More information : gerard.dray@mines-ales.fr, anne-lise.courbis@mines-ales.fr

Administrative Information

Institution: **IMT Mines Alès** (Ecole nationale supérieure des mines d'Alès), **France**
Laboratory: **Euromov Digital Health in Motion**, Univ. Montpellier, IMT Mines Ales, France.
https://dhm.euromov.eu/
Doctoral school: **I2S, Univ. Montpellier** http://www.edi2s.univ-montp2.fr/
Speciality : Computer Science

Funding: IMT Mines Alès contract
Contract duration: 36 months - Trial period: 2 months
Gross monthly salary: 2 272,50 euros
Working time: full time (151.67 hours / month)

Supervising: Supervising is done by lecturer/researcher from IMT mines Ales, France.
Director: Gérard Dray (Prof.) Co-supervisors: Anne-Lise Courbis (Ass. Prof.), Binbin Xu (Ass. Prof.), Thomas Lambolais (Ass. Prof.)

Beginning: Oct 2024

Languages: French or English

The Institution and the Host Laboratory

The **Institute Mines-Télécom (IMT)** is a French public establishment dedicated to higher education and innovative research and, as it represents the ministries of industry and digital technology, it is the largest group of engineering schools in France. The IMT brings 11 public engineering across France together. Collectively, they train 13,500 engineers and doctoral students as well as employing 4,500 women and men and manage a budget of €400M within 55 research centers connected to the schools. IMT publishes 2000 publications each year and registers 60 patents.

IMT Mines Alès is one of the schools of IMT. With its 175 years of history of service to science and industry, the school employs 350 people and trains more than 1100 students, engineers and researchers. Its three teaching and research poles work in the areas of risk environment, materials, civil engineering, industrial engineering and digital technology. The values promoted at the school are boldness, commitment, sharing and excellence. The school spurs on job mobility projects.

Research Unit "EuroMov Digital Health in Motion" is a new research unit that was officially inaugurated in January 2021. This research collaboration involves the French institutions IMT Mines Alès and the University of Montpellier in partnership with the university hospitals of Montpellier and Nîmes. The research scope promotes cross-fertilization across three main domains of artificial intelligence, movement sciences and health. The research aims to understand the behavioural plasticity of humans in order to consider new therapeutic approaches and improve sensorimotor recovery, whilst providing a platform for innovation of new digital approaches.

The main objective of study of the EuroMov Digital Health in Motion concerns human and digital plasticity seen through the prism of human movement. Human plasticity or neuroplasticity refers to the brain's ability to evolve and adapt throughout life and specific conditions. In addition to genetic factors and the environment in which a person evolves, a subject's actions and movements play a determining role in brain plasticity. Understanding the dynamic brain-movement relationships at different levels and scales will allow to promote brain plasticity and in turn improve sensorimotor recovery. The analysis of the mechanisms underlying neuroplasticity will aid, by analogy or mimicry, the development of new models for machine learning alongside the adaptive control of complex systems, to better manage human machine interaction, and the application of sensitive software systems.