Meet-up 2019 | Doctorants & Industrie

BENEFITS OF SEMI-SUPERVISED LEARNING TECHNIQUES IN RECOVERING TRACEABILITY LINKS BETWEEN DESIGN ARTIFACTS

Emma EFFA[1,2]

Marie-Pierre GERVAIS[2], Reda BENDRAOU[2], Stephen CREFF[1]

[1] IRT SystemX [2] LIP6-MoVe

Abstract

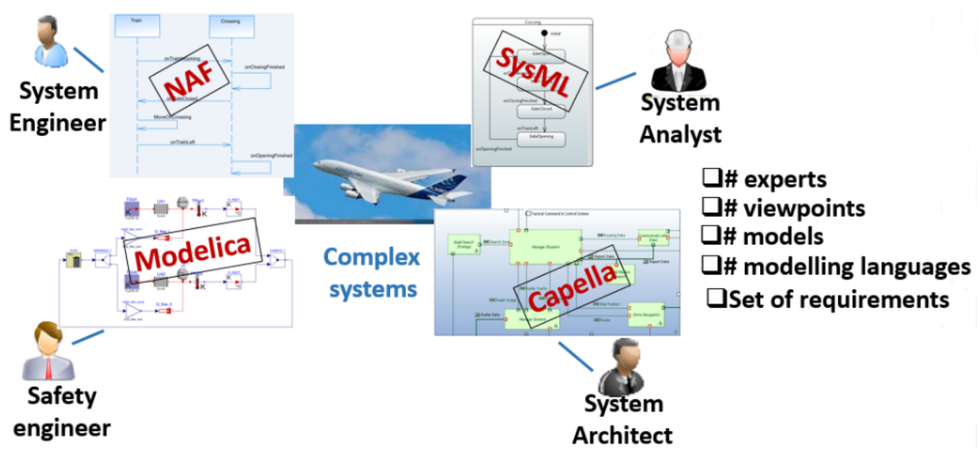
During the development of complex systems, several enterprises exchange a large number of heterogeneous models and requirements. All along the system's life cycle, these artifacts, which are linked to each other and produced in different modeling tools, are constantly evolving. In such environment, it is necessary to manage the impact of the different changes occurring in the different design spaces. Traceability management meets this need.

However, establishing links between requirements and models in complex systems engineering requires dealing with a large volume of artifacts. For example, a specification of an autonomous vehicle with 3,000 requirements and 400 model elements, it would theoretically be necessary to check about one million of potential links. Although several approaches have been proposed for identifying traceability links, the validation process is always time-consuming and error-prone.

In this thesis, we propose a semi-supervised approach that learns through a probabilistic model to recognize links or non-links from similarity measures and scores. The thesis investigates the benefits of Information Retrieval (IR) techniques and the latest advances in Natural Language Processing (NLP) ones to suggest stakeholders with candidate semantic links. This approach, implemented through the *Aggregation Trace Links Support* (ATLaS) framework, provides a quantitative confidence measure on each candidate link. This measure allows the expert in the validation phase to optimize his verification efforts while reducing the risks of error.

1 CONTEXT

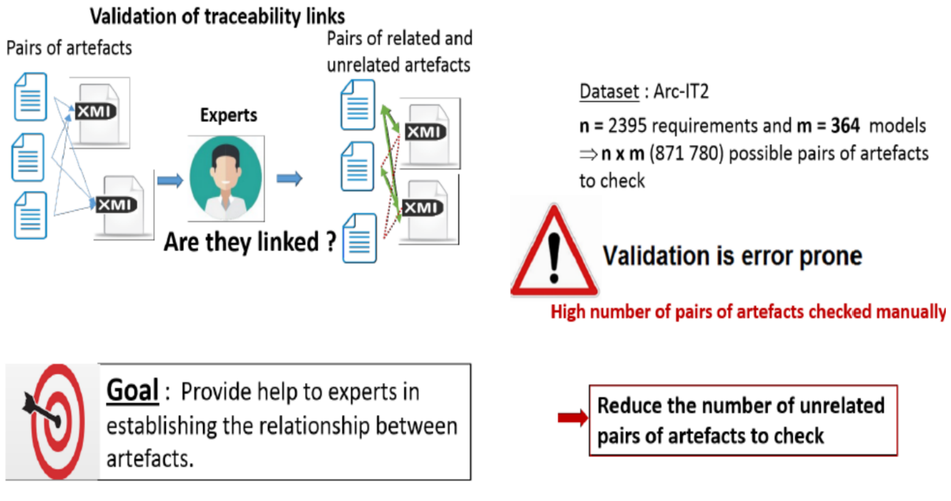
- Collaborative engineering in an Extended Enterprise context.



Multiple teams; different cultures; different tools; different formalisms…

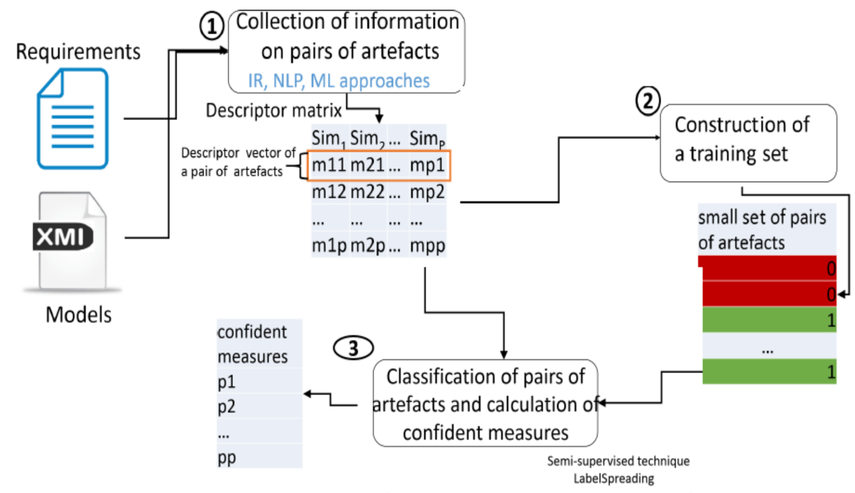
2 CHALLENGES

- How to decide whether a link exists or not between two given artifacts ?



3 CONTRIBUTIONS

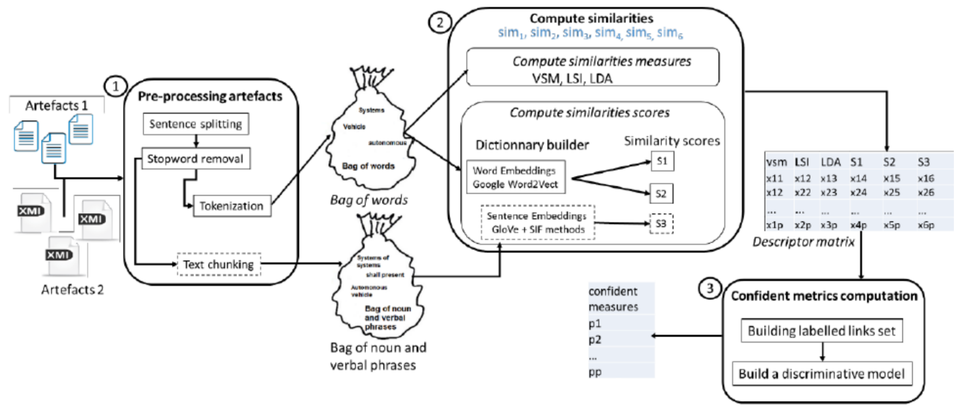
- A semi-supervised learning approach based on Information Retrieval and Natural Language Processing techniques.



Use of VSM, LSI, LDA, Word Embeddings, Sentence Embeddings, Labelspreading.

4 RESULTS

- The *Aggregation Trace Links Support* (ATLaS) framework.



Experiments made on industrial and academic case studies.

5 FUTURE WORK

- Towards the capture of specific semantic
  - Evaluate the benefits of a specific domain dictionary
- Towards improving the quality of the training set
  - Improvement of the heuristics by using related pairs of artifacts from existing projects (i.e. transfert learning) ;
  - Spatial distribution of pairs of artefacts in the descriptor space.

References

[1] Emma Effa Bella, Stephen Creff, Marie-Pierre Gervais, and Reda Bendraou. Atlas: A framework for traceability links recovery combining information retrieval and semi-supervised techniques. In *Proceedings of the 23rd IEEE International Conference In the Enterprise Computing Conference (EDOC)*, 2019.

[2] Emma Effa Bella, Marie-Pierre Gervais, Reda Bendraou, Laurent Wouters, and Koudri Ali. Semi-supervised approach for recovering traceability links in complex systems. In *Proceedings of the 23rd IEEE International Conference In the Engineering of Complex Computer Systems (ICECCS)*, 2018.

Acknowledgements

This research work has been carried out in the framework of the Technological Research Institute SystemX, and therefore granted with public funds within the scope of the French Program "Investissements d'Avenir".

Sorbonne UniversitÃ‌'s - Campus Marie-Pierre Curie
LIP 6
Emma EFFA- emma.effa-bella@lip6.fr