

Analyse d'exigences systèmes dans le projet UNSEL

Aurélien Lamercerie*, David Rouquet**,
Valérie Bellynck*, Christian Boitet***, Vincent Berment****,***

* G-INP/LIG/GETALP, ** Tétrás-Libre, *** UGA/LIG/GETALP, **** CS Group

Cet article est un résumé de 2 pages accompagnant un poster présenté à la conférence EGC-2022. Une version étendue en 12 pages est disponible sur HAL¹.

Le projet UNSEL² vise à fournir des outils pour accompagner la spécification de "systèmes de systèmes" (par exemple, un système de communications sol-air pour un aéroport). Un tel système est construit à partir d'un cahier des charges, puis de documents de plus en plus détaillés. Une spécification est un ensemble structuré d'exigences, qui sont majoritairement des énoncés descriptifs ou prescriptifs en langue naturelle (LN).

Les problèmes liés aux spécifications sont *l'incohérence, l'incomplétude et l'inadéquation*³. De plus il faut parfois les traduire, si possible avec *sens garanti*, pour des équipes délocalisées. Ces problèmes, dont l'enjeu est important, sont actuellement peu ou mal traités au niveau opérationnel. Voici l'approche proposée (et implémentée) pour les résoudre.

1. Production pour chaque exigence d'une structure-mmc (multiple, multiniveau et concrète) factorisant les analyses linguistiques possibles, grâce à l'environnement ARIANE-H.
2. Désambiguïsation interactive et intuitive en langue source, produisant une structure-umc (unique, i.e. désambiguïsée), dont l'énoncé est le "mot des feuilles".
3. Transformation en une structure-uma abstraite (transformation de certains nœuds en traits, par exemple la détermination pour les articles).
4. Transformation en un graphe UNL qui est en fait la représentation d'une structure abstraite d'un énoncé anglais équivalent dont les "mot" sont remplacés par des "lexèmes interlingues" (UW) dénotant leurs sens (déjà désambiguïsés).
5. "Extraction sémantique" de "factoïdes" et recherche d'anomalies au niveau d'un sous-ensemble d'exigences, ce qui peut déclencher un "dialogue de compréhension" (en LN bien sûr). On peut parler de "co-construction du sens".
6. Enrichissement de l'ontologie "métier" courante, initialement réduite à une petite ontologie "cadre". C'est un aspect très novateur.
7. Retour à une ou plusieurs formulations en LN (par exemple, en langage contrôlé ou simplifié, dans la langue source ou dans une autre).

On représente le sens d'un énoncé (français par exemple) sous forme d'un (hyper)-graphe UNL, qui est une structure sémantique abstraite d'un énoncé anglais équivalent (voir figure 1). Les symboles de relations et d'attributs sémantiques proviennent de l'anglais, ainsi que les

1. A. Lamercerie *et al.* Analyse d'exigences systèmes dans le projet UNsel. 2021. (hal-03475408)
2. UNsel (pour Universal Networking system engineering Language) est un projet RAPID, financé par la DGA.
3. J. Dick *et al.* Requirements Engineering. Springer, 2017.

"lexèmes interlingues" (UW), de sorte que les graphes UNL sont compréhensibles et constructibles par tous les chercheurs et développeurs du monde. Un UW est composé d'un mot-vedette (*headword*) et d'une liste de restrictions sémantiques, ce qui permet de dénoter des acceptions n'existant pas en anglais (e.g. 'tatami', ou certains niveaux de politesse, ou 'amerrier').

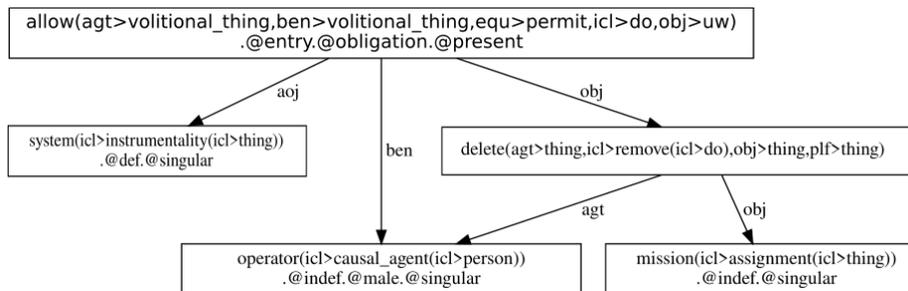


FIG. 1 – Graphe UNL pour “Le système doit permettre à un opérateur de supprimer une mission”.

Afin d'obtenir une chaîne d'extraction entièrement fondée sur les standards du Web sémantique du W3C (RDF, OWL, SPARQL, SHACL), nous avons développé dans le cadre du projet un convertisseur des graphes UNL vers une sérialisation dite *UNL-RDF*. Il est disponible sous forme d'un exécutable Java⁴ et d'un service Web⁵. Le processus d'extraction de contenu utilise une technique d'analyse par transduction sémantique adaptée de la thèse du premier auteur. On part d'un graphe UNL, que l'on transforme et enrichit, produisant une nouvelle structure formelle. Cette étape permet la construction automatique d'une ontologie OWL à partir d'une ontologie cadre passée en paramètre. L'approche UNSEL a été implémentée et appliquée à un sous-ensemble de 40 exigences⁶ tirées d'un corpus d'exigences système réel fourni par la DGA (*SRSA-IP*), composé de 367 exigences décrivant un système de communication sol-air. Les résultats sont très encourageants (voir démonstration à ECG-2022).

Summary

This paper presents the application of a semantic content extraction method in an industrial context, with the objective of automatic verification of system requirements written in natural language. The extraction step uses a semantic transduction analysis, implemented using the W3C Semantic Web standards (RDF, SPARQL, SHACL). A linguistic representation of the texts, in the form of UNL (Universal Networking Language) graphs, is exploited to provide a semi-formal structure independent of the source language. The developed tools then build an OWL ontology from the system specifications, expressed by unconstrained statements. Finally, an automatic verification of the requirements is performed using generic SPARQL rules and logical reasoners. A practical implementation on requirements from real documents has produced encouraging results.

4. <https://gitlab.tetras-libre.fr/unl/unlTools>

5. <https://unl.demo.tetras-libre.fr/>

6. sélectionnées dans *SRSA-IP* pour leur complexité linguistique et leur représentativité du corpus