

Fouille de Textes - Text Mine

ARES : un extracteur d'exigences pour la modélisation de systèmes

Aurélien Lamerterie

28 janvier 2020

① Introduction

- Système Cyber-Physique
- Modèles pour la conception de système
- Proposition : extraction des exigences d'un cahier des charges

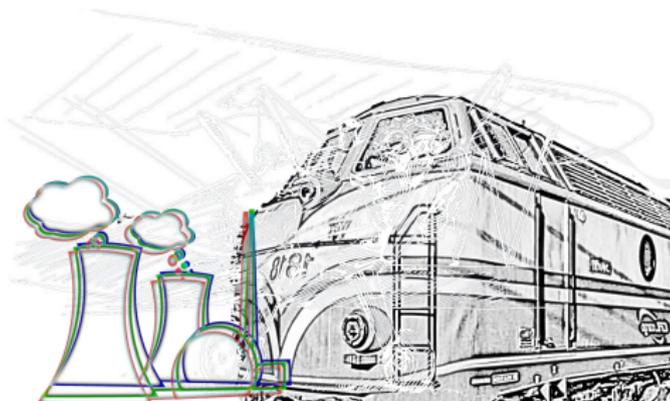
② Méthode d'extraction d'exigences en deux phases

- Analyse syntaxico-sémantique (Textes vers AMR)
- Transformation sémantique (AMR vers Règles Abstraites)

③ Conclusion et perspectives

Objectif : Apporter une aide à la conception de systèmes cyberphysiques

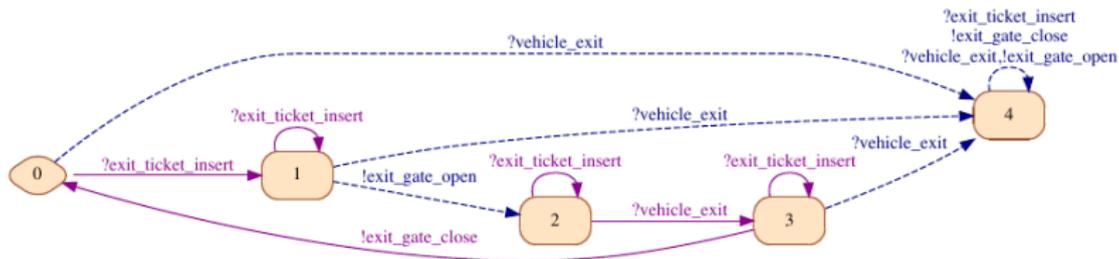
- Système Cyber-Physique :
 - logiciels en interaction avec des entités physiques
- Exemples : avion, centrale d'énergie, réseau ferroviaire...
- Enjeux : outiller la vérification des exigences ; détecter des exigences incohérentes, redondantes ou incomplètes



- Cadre : méthode formelle reliant les comportements possibles ou interdits d'un système
- Besoin d'un formalisme adapté
 - Automate \rightarrow pas de conjonction, pas de quotient
 - LTL / CTL¹ \rightarrow pas de composition, pas de quotient
 - mu-calcul² \rightarrow complexité de calcul élevé
- Automate Modal³ \rightarrow bon compromis, sans propriété d'états
- DPAA⁴ \rightarrow bon compromis, avec propriétés d'états

-
1. "Synthesis ... for ... time temporal logic", Clarke, Emerson 1981
 2. "Rudiments of mu-calculus", Arnold, Niwinski 2001
 3. "A Modal Interface Theory for Component-based Design", Raclet 2011
 4. "Une algèbre des DPAA ...", Lamerцерie 2019

Ex. : modèle représentant le comportement d'un parking de voiture



1. "Mica : A Modal Interface Compositional Analysis Library", Caillaud 2011.
2. La barrière d'entrée doit s'ouvrir une fois le ticket émis.

Proposition : ARES

- 1 Logiciel (prototype) permettant d'extraire les exigences en partant d'énoncé en langage naturel
- 2 Données traitées : une liste d'énoncés en anglais
- 3 Représentation intermédiaire : AMR¹ pour chaque énoncé
- 4 Données en sorties : listes d'entités, de propriétés et de définitions abstraites d'exigence (règles comportementales)

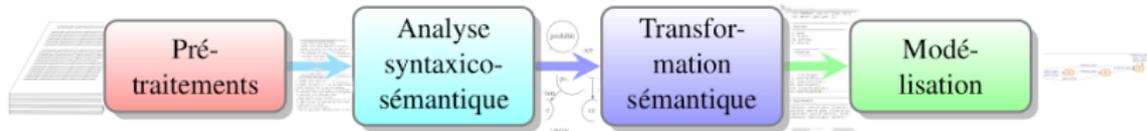


Figure – Chaîne de traitements du logiciel ARES

1. "Abstract Meaning Representation for Sembanking", Banarescu 2013

- 1 Introduction
 - Système Cyber-Physique
 - Modèles pour la conception de système
 - Proposition : extraction des exigences d'un cahier des charges
- 2 **Méthode d'extraction d'exigences en deux phases**
 - Analyse syntaxico-sémantique (Textes vers AMR)
 - Transformation sémantique (AMR vers Règles Abstraites)
- 3 Conclusion et perspectives

Exemple : cahier des charges d'un système d'accès pour un parking de voitures

R_1 The passage of a car is prohibited if the security gate is closed¹.

R_2 Entry gate must open once a ticket has been issued².

R_j ...

1. Le passage d'une voiture est interdit si la barrière de sécurité est en position fermée.

2. La barrière d'entrée doit s'ouvrir une fois le ticket émis.

Abstract Meaning Representation

- Formalisme d'annotation sémantique pour les textes en langage naturel
- Introduit par Banarescu et al (2013)
- Compact, lisible, annotation sémantique complète pour chaque phrase
- Simple en gardant un bon niveau d'expressivité

Extraction d'exigences

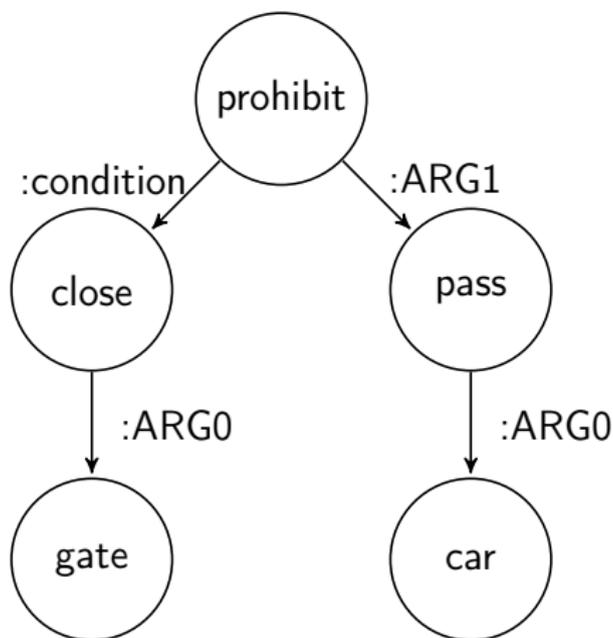
Phase 1 : dérivation syntaxico-sémantique

- Identification des entités,
- Basé sur les rôles sémantiques du corpus PropBank¹,
- Prise en compte de nombreux phénomènes linguistiques (modalités, négation, condition, conjonction ...)
- Parseur utilisé : CAMR²

1. "From Treebank to PropBank", Kingsbury and Palmer, 2002

2. "A transition-based algorithm for AMR parsing", Wang et al, 2015

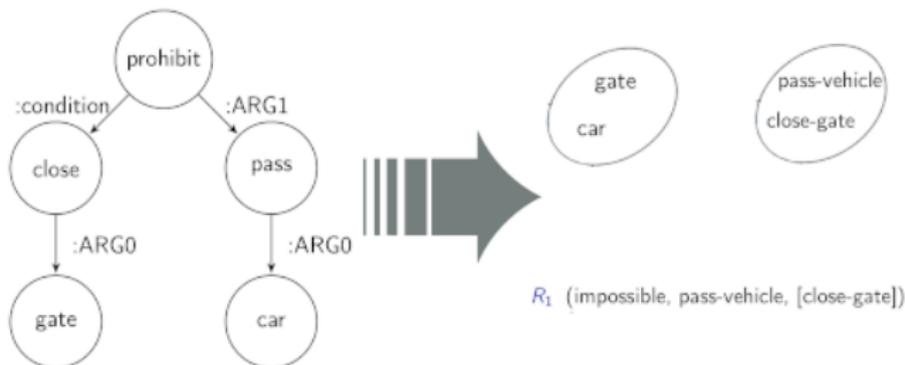
(Exemple) AMR correspondant à l'exigence R_1 ¹



1. The passage of a car is prohibited if the security gate is closed.

Objectif :

- extraire les entités et les propriétés
- transformation de la représentation sémantique
- définition formelle attendue : relation entre un contexte et une propriété associée à une modalité déontique (interdiction, possibilité, nécessité)



Algorithme de transformation sémantique (principes)

- 1 Catégorisation des instances (noeuds) du graphe AMR
- 2 Évaluation (logique) du graphe AMR
- 3 Extraction des exigences correspondant à un motif sémantique (défini sous la forme d'une formule logique)

Catégorisation des instances (noeuds) du graphe AMR

- Catégories pour les phénomènes sémantiques
 - Modality (ex. : possible, prohibit)
 - Temporality (ex. : after, before)
 - LogicConnector (ex. : and, or)
- Catégories différenciant entité et propriété
 - Entity : instance unitaire, sans argument fondamental¹
 - Property : instance avec arguments fondamentaux¹

1. argument fondamental : core-role (ARG0, ARG1...)

Évaluation (logique) du graphe AMR

- Prédicats sur les noeuds de l'AMR (ex. : *isModality*)
- Construction d'un ensemble de formules atomiques (faits) vérifiées par les noeuds de l'AMR
- ex. : *isModality(i1)*, *isProperty(i2)*, *isARG1(i1, i2)* ,
isARG1(i2, i3), *isEntity(i4)*

Définition de motifs sémantiques sous la forme d'une formule logique (lambda-calcul)

Exemples :

- $\lambda x. \lambda y. \lambda z. isModality(x) \wedge isProperty(y) \wedge isProperty(z) \wedge isARG1(x, y) \wedge isCondition(x, z)$
- $\lambda x. \lambda y. \lambda z. isModality(x) \wedge isProperty(y) \wedge isProperty(z) \wedge isDomain(x, y) \wedge isTime(x, z)$
- $\lambda x. \lambda y. \lambda t. \lambda z. isModality(x) \wedge isProperty(y) \wedge isProperty(z) \wedge isTemporality(t) \wedge isDomain(x, y) \wedge isTime(x, t) \wedge isOp1(t, z)$

(Exemple) analyse des exigences R_1 ¹ et R_2 ²

- Entités : "security-gate", "vehicle"
- Propriétés : "close-gate", "pass-vehicle", "issue-ticket", "open-gate"
- Exigences :
 - R_1 (impossible, pass-vehicle, [close-gate])
 - R_2 (necessary, open-gate, [issue-ticket])

-
1. R_1 : The passage of a car is prohibited if the security gate is closed.
 2. R_2 : Entry gate must open once a ticket has been issued

- ① Introduction
 - Système Cyber-Physique
 - Modèles pour la conception de système
 - Proposition : extraction des exigences d'un cahier des charges
- ② Méthode d'extraction d'exigences en deux phases
 - Analyse syntaxico-sémantique (Textes vers AMR)
 - Transformation sémantique (AMR vers Règles Abstraites)
- ③ Conclusion et perspectives

- ARES : logiciel pour extraire les exigences à partir d'énoncés en langage naturel (anglais)
- Processus automatisé :
 - sans contrainte particulière sur le texte
 - sans connaissance à priori du système étudié
 - sans ontologie particulière
- Expérimentation (limité) : résultat intéressant sur quelques cas de test

- Formalisation de la proposition
- Meilleure couverture des différents phénomènes linguistique / sémantique
- Validation rigoureuse sur un corpus plus complet
- Intégration complète d'une chaîne de traitement de bout-en-bout

Pour plus de détails et démonstration :

- contact mail : aurelien.lamercerie@inria.fr