

Construction et vérification de la spécification formelle d'une configuration d'environnement intelligent issue d'une description textuelle

Encadrement : Brigitte Grau (LIMSI), Ecole doctorale d'informatique de Paris 11

Co-encadrements : Catherine Dubois (CEDRIC), Yacine Ghamri (LIGM)

Allocation de recherche DIGITEO

Contexte et objectifs

Des travaux récents se sont attachés à la vérification et à la validation de réglementations: réglementations de sécurité dans les aéroports, protocoles médicaux par exemple. Dans ces travaux, les modèles formels qui servent à la vérification par preuve ou model checking sont en général élaborés directement par l'utilisateur à partir des textes.

Notre objectif ici est d'appliquer le même principe dans le cadre de la configuration d'environnements intelligents grâce à des descriptions textuelles fournies par des techniciens ou des utilisateurs avertis (fêrus de technologie). En effet au vu des possibilités offertes par les objets (capteurs, actionneurs, périphériques de contrôle) existants, une multitude de configurations sont envisageables pour un environnement intelligent. Dans le projet ENVIE VERTE, afin de tirer partie du potentiel offert par ces environnements, nous proposons d'élaborer des outils de configurations fiables et faciles à utiliser.

Nous proposons dans cette thèse de travailler sur une partie de ce processus. Plus précisément, nous proposons ici d'élaborer un modèle complet allant du texte à la représentation formelle, qui requiert une intervention de l'utilisateur uniquement quand cela s'avère nécessaire.

L'originalité de ce sujet repose sur le fait de toujours maintenir une relation entre les représentations construites et le texte, ainsi que dans une construction itérative de la spécification.

L'analyse de spécifications de systèmes décrites en langue naturelle afin d'en proposer des représentations permettant ensuite d'engendrer du code ou des spécifications formelles a été étudié par différents types d'approches, mettant toutes en évidence la difficulté de traiter la langue naturelle de part ses ambiguïtés et son imprécision. On peut distinguer deux approches :

- celles dont l'objectif est d'analyser le texte et en extraire des spécifications en un langage semi-formel, de type UML. Généralement ces travaux partent de l'hypothèse d'un langage contraint ;
- celles dont l'objectif est de fournir des outils linguistiques pour aider le concepteur à construire ses spécifications, et traiter les documents dans leur généralité.

Le travail projeté se situe dans cette deuxième approche, de manière analogue à ce qui se fait dans Ambriola et Gervasi [1], où l'utilisateur définit préalablement un glossaire permettant d'assigner aux termes du document un type sémantique, et le système propose ensuite une représentation abstraite construite par application de règles de transformations s'appuyant sur les types sémantiques reconnus dans le texte. On retrouve cette approche dans Lee & Bryant [6] qui transforment chaque phrases en une représentation intermédiaire, qui tient compte de la structure du texte pour mémoriser un contexte et du type des phrases, avant de produire une représentation en un langage formel, VDM ++. Les inférences et repérages de contradictions sont réalisés principalement sur les représentations formelles.

Objectif de la 1ère année

1. Analyse des textes (les descriptions en langage naturel des configurations et contrôles souhaités) par la réutilisation et l'adaptation d'outils de fouille de texte permettant d'annoter ceux-ci : repérage de termes du domaine sous leurs différentes formes fondé sur une utilisation d'une ontologie du domaine, qui sera éventuellement à compléter par acquisition semi-automatique de termes, repérage de définitions, repérage de passages formulant des spécifications par reconnaissance d'expressions linguistiques. Cette analyse s'appuiera sur les travaux déjà réalisés dans le cadre de la recherche d'informations précises (en question-réponse [4], [6] et en extraction d'information [7]).
2. Réalisation de différents types d'interaction avec l'utilisateur du système : proposition de

modélisation si l'on peut reconnaître les formes syntaxico-sémantiques menant à cette modélisation ou produire une analyse partielle et proposer des éléments du modèle que l'utilisateur choisira d'assembler à sa convenance. Le choix du formalisme de représentation conceptuelle sera guidé par le souci de pouvoir le présenter à l'utilisateur sous forme compréhensible, par un formalisme graphique tel qu'UML, et par le fait de pouvoir engendrer ensuite une représentation formelle de la spécification et la réalisation de la configuration projetée.

Objectif des 2 dernières années

Après avoir dégagé une représentation conceptuelle adaptée aux besoins analysés ainsi qu'aux formalismes visés, il s'agira de produire le modèle formel. On choisira d'en extraire un modèle écrit en Focal [2] complété éventuellement par d'autres annotations formelles.

Un prototype de traduction automatique sera développé au cours de la thèse. Il prendra en entrée la représentation interne ainsi que des annotations délimitant les éléments à retenir dans la spécification formelle. Ces annotations seront données par le concepteur de l'environnement intelligent.

Le fait de disposer d'une spécification formelle permet d'analyser cette spécification pour en détecter les éventuelles incohérences ainsi que les manques de celle-ci. Il s'agira d'étudier comment interpréter ce diagnostic afin de retourner au texte, et en rectifier ou compléter l'analyse, et faire ainsi évoluer le modèle de spécification. Le lien gardé entre le texte et sa représentation conceptuelle et formelle permettra de cibler les passages à analyser ou de chercher des informations dans des passages qui n'ont pas donné lieu à une interprétation. Il s'agira ici aussi d'élaborer un dialogue avec l'utilisateur permettant de valider et faire évoluer la spécification construite.

Comme l'utilisation des méthodes formelles requiert une certaine expertise que l'utilisateur ne possède pas nécessairement, cela pose des problèmes de communication qui risquent de compromettre le processus de validation. Une approche assez largement adoptée pour résoudre ce type de situation consiste à intégrer spécifications formelles et spécifications graphiques. Le doctorant étudiera les méthodes de visualisation graphique associée à la représentation interne. On s'appuiera sur les techniques et outils existants de transformation de Focal à UML ([3]).

En effet l'objectif applicatif final serait de s'assurer que la configuration de l'environnement intelligent voulu par l'utilisateur est intègre au sens des propriétés que l'environnement intelligent doit avoir. Il est à noter que le raffinement (au sens réseau du terme) et le déploiement effectif des configurations ne fait pas partie des objectifs de cette thèse. Ce dernier point sera traité par d'autres personnes dans le projet.

Références bibliographiques

- [1] Ambriola V., Gervasi V., The CIRCE approach to the systematic analysis of NL requirements, Technical report: TR-03-05, Univ. Pisa, 2003
- [2] P. Ayrault, M. Carlier, D. Delahaye, C. Dubois, D. Doligez, L. Habib, T. Hardin, M. Jaume, C. Morisset, F. Pessaux, R. Rioboo and P. Weis. Trusted Software within FoCaL. C&ESAR 2008 - Trusting Trusted Computing?, Rennes, December 2008.
- [3] D. Delahaye, J.-F. Étienne et V. Viguié Donzeau-Gouge. A Formal and Sound Transformation from Focal to UML: An Application to Airport Security Regulations . In UML and Formal Methods (UML&FM), Kitakyushu-City (Japan), October, ISSE NASA Journal, Springer, 2008.
- [4] Brigitte Grau, Anne-Laure Ligozat, Isabelle Robba, Anne Vilnat et Laura Monceaux FRASQUES: A Question-Answering System in the EQueR Evaluation Campaign *LREC 2006*, Gênes, Italie, Mai 2006.
- [5] Lee B.-S., Bryant B., Contextual Knowledge Representation for Requirements Documents in Natural language, 15th FLAIRS, 2002
- [6] Anne-Laure Ligozat, Brigitte Grau, Anne Vilnat, Isabelle Robba, Arnaud Grappy Towards an automatic validation of answers in Question Answering, 19th IEEE International Conference on Tools with Artificial Intelligence - Vol.2 (ICTAI 2007) pp. 444-447, 2007
- [7] Anne-Lyse Minard, Anne-Laure Ligozat, Brigitte Grau. Extraction de résultats expérimentaux d'articles scientifiques pour le peuplement d'une base de données. 10th International Conference on statistical analysis of textual data (JADT), 2010