

Proposition de stage niveau M2

Titre : détection de formes d'incohérence dans les spécifications techniques

Mots clés : *traitement automatique des langues naturelles, raisonnement, règles métier*

Proposant : Patrick Saint-Dizier, stdizier@irit.fr

Ce stage s'inscrit dans le cadre du projet LELIE, un environnement d'aide à la rédaction de spécifications techniques (exigences, règles métier, cahiers de charge, règles de sécurité, etc.). L'un des défis les plus importants est la détection de différentes formes d'incohérences dans des spécifications qui peuvent contenir plusieurs dizaines de milliers de règles. Détecter les incohérences dès leur rédaction est capital : des risques humains (accidents), écologiques, sociaux et financiers lourds peuvent ainsi être évités.

Les spécifications et règles métier concernent des domaines très variés : règlements bancaires, d'assurances, maintenance, propriétés d'un logiciel ou de tout autre objet manufacturé, règles de gestion du personnel, code du travail, etc.

Les incohérences dans les spécifications proviennent essentiellement de trois facteurs : une mauvaise analyse du phénomène spécifié, une méconnaissance du domaine (par exemple des équipements utilisés) ou des erreurs dans la rédaction. Nous nous concentrerons sur ce dernier point, ayant en général peu d'informations sur le domaine.

Le stage consistera en une identification de quelques catégories, sur une base linguistique, à partir de corpus de spécifications. Les incohérences sont d'origines diverses, détectables (facilement) sur la base de critères linguistiques seulement.

Soient deux exigences P et Q. On regarde les doublons ou incohérences (partielles ou totales) lorsque P et Q ont un recouvrement significatif (à mesurer, modulo quelques variantes morphologiques) afin de s'assurer que le 'contexte' est le même. Certaines variantes morpho (ex. quantification ou les temps verbaux) sont significatives d'une incohérence, d'autres non. Le stagiaire utilisera les facilités d'analyse de LELIE. Le travail se fera sur le français et/ou l'anglais. Cela inclut des expressions évaluatives mal formées : P : chauffer la sonde X30 au max à 30 degrés/ chauffer la sonde X30 à au moins 50 degrés, le traitement de la négation : P1 : inclure la notion de communauté dans D / Q1 : la notion de communauté n'est pas pertinente dans D, modaux incohérents : P : it is recommended to stop the engine before E / Q : the engine shall be stopped before E, listes incomplètes ; P : a rule includes facts and its criticity./ Q : a rule includes facts and a description, etc.

Outre une analyse et une catégorisation de formes d'incohérence, le travail consistera en la modélisation puis la mise en œuvre de l'analyse des formes d'incohérences relevées. Enfin, un travail d'évaluation des résultats sera conduit sur de vrais documents industriels. Le problème de l'innocuité sera étudié (incohérences faibles qui ne gênent malgré tout pas la compréhension) ainsi que le bruit et le silence générés (détectations incorrectes d'incohérences ou oublis). Le résultat sera intégré comme Plug In sous Word et Excel indépendamment du stage mais sera utile visuellement pour l'évaluation.