

Proposition de présentation du CATI DIISCICO au séminaire INRAE Semantic Linked Data du 11 au 14 octobre

Un outil d'aide à la conception d'itinéraires techniques basé sur les ontologies PO2, naRyQ et les PRM: application aux emballages alimentaires composites

Mélanie Münch^{1,2}, Hélène Angellier-Coussy², Patrice Buche^{1,2}, Stéphane Dervaux^{1,3}, Liliana Ibanescu^{1,3}, Cristina Manfredotti³, Claire Mayer-Laigle², Magalie Weber⁴

1-CATI DIISCICO, 2-UMR IATE, Montpellier, 3-UMR MIA Paris, 4-UR BIA, Nantes

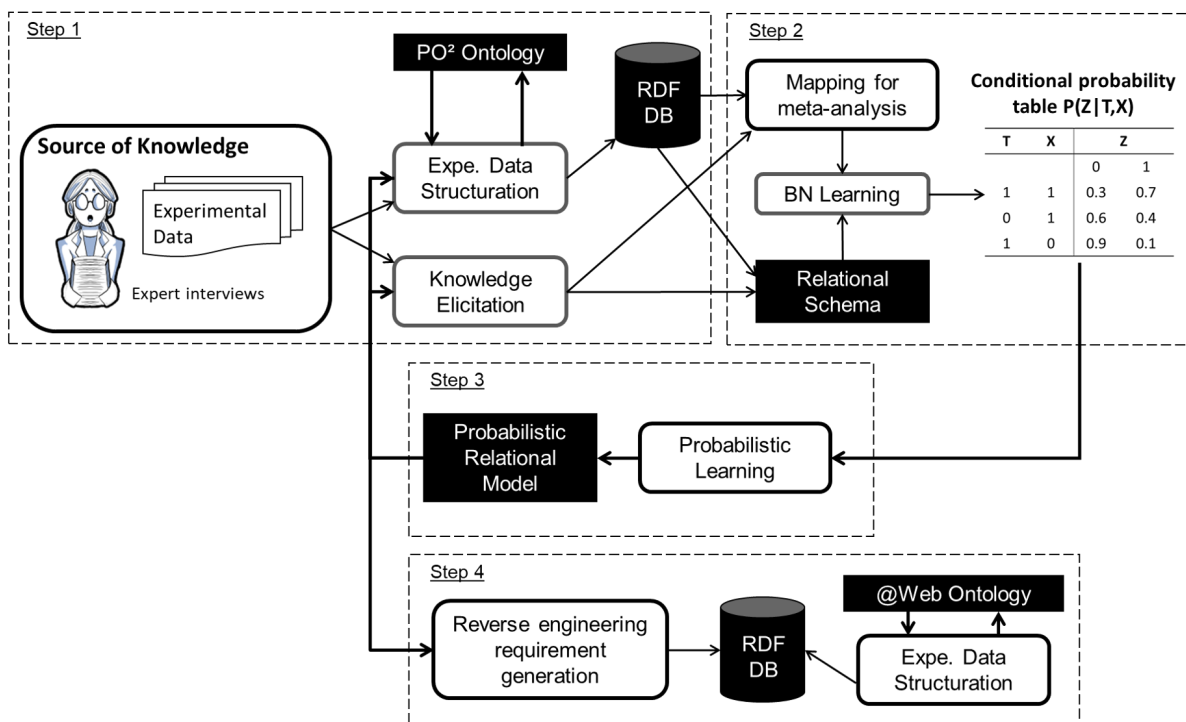


Fig. 1 - Architecture de l'outil d'aide à la décision

Résumé: Les données expérimentales produites lors de travaux de recherche pour concevoir de nouveaux itinéraires techniques de transformation de la biomasse en bio-produits sont complexes à obtenir. Afin de pouvoir tirer des conclusions à partir des données obtenues dans plusieurs projets, nous proposons un outil d'aide à la conception qui repose sur un pipeline composé de quatre étapes, illustrées dans la figure 1:

1. Création d'une base de données RDF: l'ontologie PO² (Process and Observation Ontology) est utilisée pour structurer dans une base de données RDF, appelée Bio PO²-RDF DB, des données expérimentales hétérogènes obtenues dans plusieurs projets. Bio PO²-RDF DB enregistre les caractéristiques des itinéraires techniques de transformation de la biomasse en bio-produits et les caractéristiques/propriétés des entrées et des sorties de ces itinéraires.
2. Création d'un schéma relationnel RS(Relational Schema): Le contenu de la base de données Bio PO²-RDF DB et des connaissances expertes sont combinés pour élaborer un schéma relationnel utilisé pour modéliser toutes les contraintes d'apprentissage du Réseau Bayésien (BN) qui sera construit dans l'étape suivante.
3. Apprentissage du Réseau Bayésien BN: le BN est appris à partir du RS construit lors de l'étape 2 (RS), le résultat étant un PRM (Probabilistic Relational Model)

4. Ingénierie inverse: les résultats produits grâce au PRM permettent à partir de valeurs ciblées pour les caractéristiques du produit obtenu de déterminer les valeurs optimales associées aux caractéristiques des biomasses en entrée de l'itinéraire technique. La disponibilité d'une deuxième base de données RDF structurée par le format de l'ontologie naRyQ, incluant des données de caractérisation de biomasses, permet de présenter à l'utilisateur des données intégrées des deux BD, rapprochées par les inférences du PRM. Cette étape est réalisée grâce à l'utilisation de requêtes fédérées SPARQL, sur les deux SPARQL endpoints.

Perspectives: La mise en relation de données RDF annotées avec des ontologies de domaine issues de projets de recherche différents au sein du département INRAE Transform pose la question de la constitution d'une ressource unique et homogène pour faciliter le rapprochement des données. Une réflexion est en cours afin de définir une méthode d'intégration dans une ontologie de domaine unique correspondant au périmètre du département Transform des micro-ontologies de domaine réalisées au sein des différents projets.