Proposition de présentation du CATI DIISCICO au séminaireINRAE Semantic Linked Data du 11 au 14 octobre Towards decision support system for the agri-food sector using heterogeneous scientific data annotated with an ontology and Bayesian networks: a proof of concept applied to milk microfiltration

Cédric Baudrit, P. Buche, Nadine Leconte, Christophe Fernandez, Maellis Belna, Geneviève Gésan-Guiziou

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

Towards decision support system for the agri-food sector using heterogeneous scientific data annotated with an ontology and Bayesian networks: a proof of concept applied to milk microfiltration

Cedric Baudrit[a,e], Patrice Buche[b,e], Nadine Leconte[c], Christophe Fernandez[a,e], Maëllis Belna [a,c,d], Geneviève Gésan-Guiziou[c]

a INRAE, Université de Bordeaux-I2M, 33400 Talence, France
b IATE, Univ Montpellier, INRAE, Institut Agro, Montpellier, INRIA GraphiK, France
c STLO, INRAE, Institut Agro, 35000, Rennes, France
d Boccard, Research and Development, 35360 Montauban-de-Bretagne, France
e CATI INRAE DIISCICO

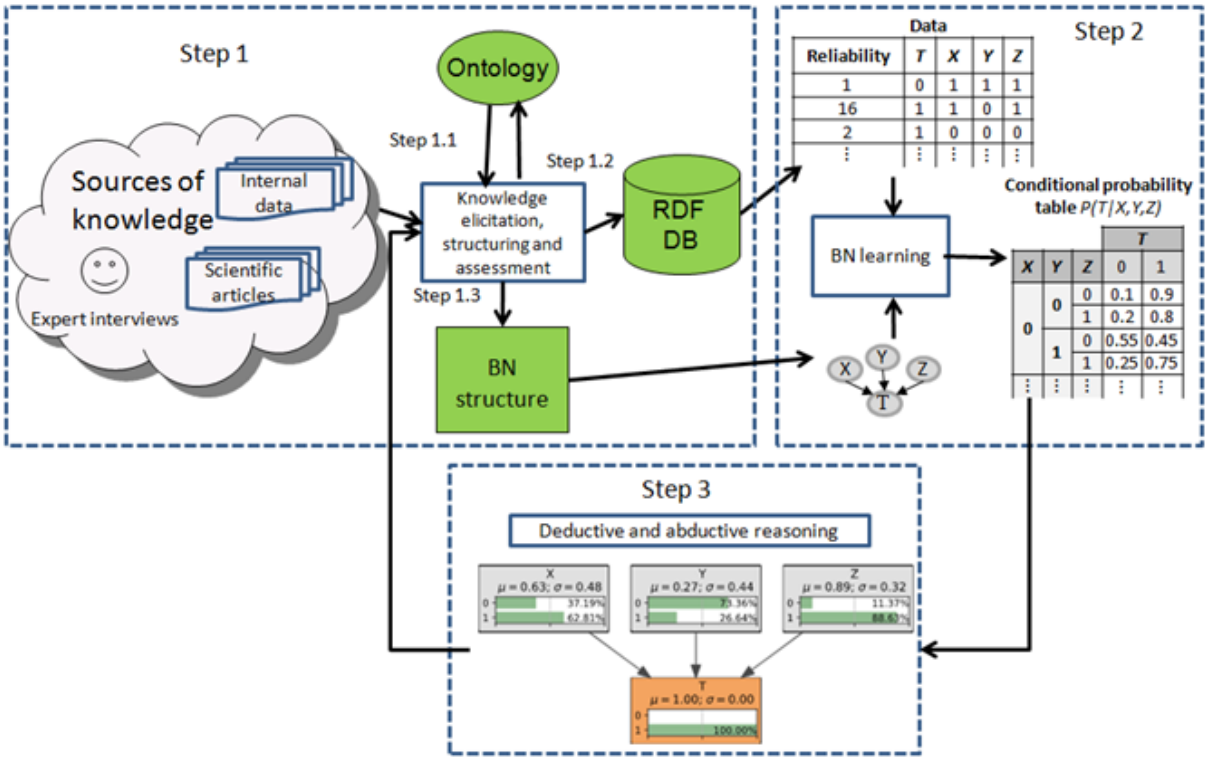

Fig. 1 - Architecture de l'outil d'aide à la décision

Résumé: With more than 9000 scientific articles published each day, the scientific literature, including experimental data and knowledge, is a valuable source of information for developing predictive models to design decision support systems. However, scientific data are heterogeneously structured, found mainly in text format and expressed using different vocabularies. This study developed a practical and versatile pipeline that combines ontology, databases and computer calculation tools based on the theory of belief functions and Bayesian networks. The objective was to design an iterative process that can structure a domain of knowledge, use domain ontology to integrate data in structured databases, and build a predictive model from ontology and data. The ontology paradigm is used to help integrate data from heterogeneous sources and build the structure of the Bayesian

network. The parameters of the Bayesian network are estimated using the integrated data taking into account their reliability. The pipeline can be used iteratively to enrich the database and the model with new data and knowledge over time without damaging the entire system. The pipeline was assessed by applying it to a complex food engineering process: skimmed milk microfiltration. , which is one of the first operations that can be carried out after harvesting. This application enabled (1) building the domain ontology of milk microfiltration to annotate state-of-the-art literature sources, (2) creating a structured database and (3) developing a predictive model of the permeation flux associated with operating conditions.