

A Data Warehouse that Gathers Several Formalisms to Capture Data Heterogeneity and Incompleteness in the Field of Food Microbiological Safety

Patrice Buche*, Juliette Dibie-Barthélemy**
Ollivier Haemmerlé**, Rallou Thomopoulos***

*INRA - Mét@risk, 16 rue Claude Bernard, F-75231 Paris Cedex 05, France

Patrice.Buche@inapg.fr,

**UMR INA P-G/INRA BIA, 16 rue Claude Bernard, F-75231 Paris Cedex 05, France
{Juliette.Dibie,Ollivier.Haemmerle}@inapg.fr

***INRA - UMR IATE - bat. 31, 2 Place Viala 34060 Montpellier Cedex 1
rallou@ensam.inra.fr

Résumé. Notre travail porte sur la conception d'un entrepôt de données dans le domaine de l'analyse du risque alimentaire. Les données expérimentales stockées dans cet entrepôt ont trois caractéristiques principales qui ont guidé la conception de l'entrepôt : elles sont hétérogènes, elles peuvent être imprécises, et l'entrepôt est par nature incomplet. Nous proposons d'utiliser trois modèles de données : le modèle relationnel, le modèle des graphes conceptuels et le modèle XML. Ces modèles ont été étendus afin de pouvoir représenter des données imprécises sous la forme de distributions de possibilités. Les bases de données construites sur ces modèles sont interrogées simultanément avec le langage MIEL++. Dans ce langage, les préférences de l'utilisateur sont représentées par des sous-ensembles flous. Des techniques de mise en correspondance floue sont utilisées pour comparer préférences et données imprécises.

1 Introduction

Since 1999, our team has been working with industrial and academic partners on several projects which concern knowledge representation in the field of predictive microbiology. In the Sym'Previous (<http://www.symprevious.org>) and e.dot projects (<http://www-rocq.inria.fr/verso/edot/>), we work on the building of a data warehouse composed of data concerning the behaviour of pathogenic germs in food products. Those data are designed to be used in a tool dedicated to researchers in microbiology or to industrials. The goal is to help them in a decision support approach in order to prevent food products from contamination.

The information we have to store in our data warehouse presents several specificities. It is *weakly-structured* because information comes from heterogeneous sources (scientific literature, industrial partners...) and is still rapidly evolving as predictive microbiology is a research field. It is *imprecise* because of the complexity of the underlying biological processes, and because of the internal imprecision of the measurement tools. The data warehouse is *incomplete* by nature since the number of experiments is potentially infinite : it will never contain information about all the possible food products