SAFE-Next : Une approche systémique pour l'intégration des connaissances du domaine dans la fouille de données complexes

Walid Ben Ahmed*,***, Mounib Mekhilef*
Michel Bigand**, Yves Page***

*LGI – Laboratoire de Génie Industriel, Ecole Centrale Paris, Grande voie des Vignes 92295 Châtenay-Malabry cedex, France { walid, mekhilef@lgi.ecp.fr} http://www.lgi.ecp.fr

**Équipe de Recherche en Génie Industriel, Ecole Centrale de Lille, BP 48, 59651 Villeneuve d'Ascq cedex, France

michel.bigand@ec-lille.fr http://www.ec-lille.fr

***LAB (PSA-Renault), Laboratoire d'Accidentologie, de Biomécanique et d'études du comportement humain, 132, rue des Suisses-92000 Nanterre vves.page@lab-france.com

Résumé. L'Extraction de Connaissances de Données (ECD) est un processus itératif dont la complexité dépend de la nature des données traitées, de la nature des connaissances à extraire ainsi que du domaine de l'application. Lorsque cette complexité est élevée, une forte implication de l'utilisateur est requise tout au long du processus et surtout dans la première et la dernière phase (i.e. préparation de données et interprétation des résultats). En combinant des approches d'ECD et d'Ingénierie de Connaissances (IC), nous développons une méthodologie descendante qui permet l'identification multivues des connaissances du domaine, leur formalisation sous forme de métadonnées ainsi que leur incorporation dans le processus d'ECD. Notre approche est appliquée dans le domaine d'accidentologie pour l'extraction à partir des bases de données d'accidents de la route des connaissances exploitables pour le développement des systèmes de sécurité embarqués dans les véhicules.

1. Introduction

La sécurité routière est un enjeu social, politique, économique et technologique. Malgré les derniers progrès en terme de réduction du nombre de tués sur la route en France (moins 22% en 2003 par rapport à 2002), ce nombre reste élevé (5732).

L'accidentologie est un domaine dont l'objectif est l'étude des accidents pour construire des connaissances permettant d'améliorer la sécurité routière. L'un des supports de cette connaissance est le concept de scénario type d'accident (STA) qui est défini par l'INRETS¹ comme un déroulement prototypique (prototype de déroulement) correspondant à un groupe d'accidents présentant des similitudes d'ensemble du point de vue de l'enchaînement des faits et des relations de causalité dans les différentes phases [Brenac et Fleury, 1999]. Un exemple de scénario est donné dans la figure suivante :

-

¹ INRETS : Institut National de REcherche sur les Transports et leur Sécurité.