

# Annotation temporelle des événements dans des dépêches épidémiologiques

Yann Guilbaud\*, Jean Royauté\*

\*Laboratoire d'Informatique Fondamentale de Marseille (LIF)  
UMR 6166 CNRS - Univ. de Provence - Univ. de la Méditerranée  
Parc scientifique et technologique de Luminy  
CASE 901, 163 Avenue de Luminy  
F-13288 Marseille Cedex 9  
yann.guilbaud@lif.univ-mrs.fr, jean.royaute@lif.univ-mrs.fr

**Résumé.** Nous nous intéressons dans cet article à l'annotation temporelle des événements décrits dans des dépêches issues du site de diffusion d'informations épidémiologiques PROMED MAIL. Après avoir présenté l'intérêt d'un marquage et d'un calcul temporel sur de telles données textuelles pour des épidémiologistes, nous présentons l'expérimentation que nous avons réalisée en utilisant des transducteurs et un calcul exploitant les intervalles d'Allen. Pour le marquage des événements, nous avons été amenés à caractériser le sous-langage de l'épidémiologie tel qu'il apparaît dans les dépêches. Des schémas d'extraction ont ensuite été implémentés sous forme de transducteurs. Le fait qu'il ne soit pas toujours possible de faire référer ces événements à une date précise nous amène à proposer une représentation de la temporalité de ces événements basée sur l'algèbre des intervalles temporels de Allen ainsi que sur ses extensions. Nous utilisons alors les relations décrites entre ces intervalles pour déterminer, à partir des traits temporels explicites, certains traits temporels implicites. Pour faciliter et rendre plus précis le calcul nous avons introduit des scénarii permettant l'extension des relations temporelles explicites en vue de déterminer certaines temporalités implicites. Nous proposons une application qui utilise les annotations obtenues par transducteurs pour réaliser ce traitement. Ce programme, implémenté en Java, complète et calcule les temporalités implicites. Enfin, nous présentons les résultats que nous obtenons à partir de différents corpus réalisés, représentatifs de différentes maladies.

## 1 Introduction

Les dépêches épidémiologiques sont de courts textes narratifs, de style journalistique, rapportant une succession d'événements connectés décrivant l'évolution d'une épidémie. Ces dépêches constituent une ressource précieuse pour les épidémiologistes chargés du suivi de ces maladies. Cependant sous leur forme actuelle il n'est pas possible d'envisager des traitements automatiques exploitant leur contenu informationnel. Ces textes véhiculent trois types d'informations importantes : une information factuelle qui concerne les événements en eux-mêmes (l'épidémie et ses différentes mani-