Chapter 9

MedISys: Medical Information System

Jens P. Linge
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Ralf Steinberger
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Flavio Fuart
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Stefano Bucci
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Jenya Belyaeva
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Monica Gemo
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Delilah Al-Khudhairy
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

Roman Yangarber
University of Helsinki, Department of Computer Science, Finland

Erik van der Goot
Joint Research Centre of the European Commission Institute for the Protection and Security of the Citizen Global Security and Crisis Management Unit, Italy

ABSTRACT

The Medical Information System (MedISys) is a fully automatic 24/7 public health surveillance system monitoring human and animal infectious diseases and chemical, biological, radiological and nuclear (CBRN) threats in open-source media. In this article, we explain the technology behind MedISys, de-
scribing the processing chain from the definition of news sources, scraping and grabbing articles from the internet, text mining, event extraction with the Pattern-based Understanding and Learning System (PULS, developed by the University of Helsinki), news clustering and alerting, to the display of results. The web interface and service applications are shown from a user’s perspective. Users can display world maps in which event locations are highlighted as well as statistics on the reporting about diseases, countries and combinations thereof and can apply filters for language, disease or location or filters with orthogonal categories, e.g. outbreaks, via their browser. Specific entities such as persons, organizations and locations are identified automatically.

INTRODUCTION

In many fields, professionals need to scan vast quantities of information from multiple sources on a daily basis, e.g. journalists have to keep up with incoming news stories, press officers need to react quickly to evolving stories, and investors follow the latest developments affecting the stock markets.

In the area of Public Health (PH), national and international authorities continually monitor the widest possible set of available sources of information. In their daily surveillance routine, public health authorities use indicator-based and event-based surveillance tools to identify evolving public health threats and to track ongoing incidents [Paquet et al. 2006]. A broad range of threats needs to be covered, from outbreaks of communicable diseases, terrorism cases such as the deliberate release of biological or chemical agents, contaminations of food and feed to chemical or nuclear incidents. Public health authorities employ experts in these domains who monitor all available sources of information. Timely monitoring is critical for the risk assessment.

Indicator-based surveillance systems collect structured data from health-care centres, clinicians, etc., and propagate it through official channels to the top-level authorities. Event-based surveillance refers to the activity of monitoring a much wider range of unstructured sources to detect actual or perceived threats to public health. Event-based monitoring within public health authorities is becoming increasingly focused on event detection using informal media sources (news wires, online newspapers, specialist blogs, etc.) on the internet.

Event-based media-monitoring systems can greatly facilitate the work of the analysts by identifying potentially relevant news items [Steinberger et al. 2008]. These systems can be broadly classified in news aggregators, automatic systems and moderated systems [Linge et al. 2009]. News aggregators collect articles from several sources, usually filtered by language or country. Most news aggregators simply retrieve RSS (Really Simple Syndication) feeds from news providers, e.g. online versions of newspapers. Automatic systems such as MedISys, HealthMap [Freifeld et al. 2008] and BioCaster Global Health Monitor [Collier et al. 2008] collect and automatically analyse articles to facilitate the work of the analysts. Automatic systems differ in the range of information sources, their language coverage, the speed of delivering information, level of analysis and visualization.

Moderated systems rely on dedicated teams of human analysts who manually scan and analyse the retrieved documents. GPHIN (Canada) [Mykhaylovskiy et al. 2006], ProMED-Mail [Madoff 2004], and Argus [Wilson et al. 2008] (USA) are examples of moderated systems.

The trade-offs between automatic processing and human moderation lie in speed, accuracy of analysis, and cost. Human moderation is resource-intensive; GPHIN and Argus employ dozens of highly-qualified (and highly-paid) analysts to