

Scanning the Emerging Infectious Diseases Horizon - Visualizing ProMED Emails Using EpiSPIDER

Herman Tolentino MD¹, Raoul Kamadjeu MD, MPH¹, Paul Fontelo MD, MPH², Fang Liu MS², Michael Matters PhD, MPH¹, Marjorie Pollack MD³, Larry Madoff MD^{3,4}

¹ Public Health Informatics Fellowship Program, Office of Workforce and Career Development, Centers for Disease Control and Prevention, Atlanta GA 30333, ² Office of High Performance Computing and Communications, National Library of Medicine, 8600 Rockville Pike, Bethesda, Maryland 20894, ³ ProMED Mail (<http://www.promedmail.org>), ⁴ International Society for Infectious Diseases, Channing Laboratory and Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School

Background. The Internet has created an information revolution that spans across all knowledge domains and removes temporal and geographic barriers. Various disparate tools allow individuals to communicate with each other across these barriers. We also have an abundance of electronic resources containing health information locked inside free text components. The lack of integration of these tools and electronic resources has prevented harnessing of information for use in integrated and novel ways. We developed an application for "Semantic Processing and Integration of Distributed Electronic Resources for Epidemiology" or EpiSPIDER (<http://www.epispider.org>), an integrative web-based information processing system that uses these tools and electronic resources to create an information environment for enhancing the surveillance of emerging

information from news syndication sites like Reuters. We extract location information from these sources using natural language processing and geocode them using the Yahoo and Google geocoding services. Figure 1 shows the main inputs, processing and outputs of EpiSPIDER. We collaborated with the National Library of Medicine to link the topics extracted from ProMED reports to relevant PubMed articles using askMEDLINE, a natural language query system [2]. Lastly, we generate filtered and unfiltered RSS and georSS (geocoded RSS) feeds containing summaries of daily ProMED reports and make them available through the EpiSPIDER web site.

Results. EpiSPIDER began collecting and processing ProMED reports on week 13 of 2006. Since then, in collaboration with ProMED staff, we have refined EpiSPIDER algorithms that recognize location entities from free text. Using both the Google Maps and Yahoo Maps APIs, EpiSPIDER now generates country-level maps for all countries covered by ProMED reporting, as well as state-level maps for the US, Canada and a few selected countries. EpiSPIDER reconstructs events before week 13 through recursive spidering and also links mapped topics to PubMed citations. As a data provider, it makes available filtered and unfiltered RSS feeds for downloading by agencies and organizations from Atlanta, Georgia, the North Carolina Research Triangle and Washington DC.

Conclusions. We have developed EpiSPIDER as a tool that integrates distributed electronic resources and that we can extend and improve to provide decision support for managing the global threat from emerging infectious diseases. Future work includes refinements in the NLP algorithms that parse news wire feeds and mapping them to ProMED topics, mapping topics to Unified Medical Language System (UMLS) concepts, and email-based data and information exchange to and from EpiSPIDER.

References

1. Madoff LC, Woodall JP. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch Med Res.* 2005 Nov-Dec;36(6):724-30. Review.
2. Fontelo P, Liu F, Ackerman M. askMEDLINE: a free-text, natural language query tool for MEDLINE/PubMed. *BMC Med Inform Decis Mak.* 2005 Mar 10;5(1):5

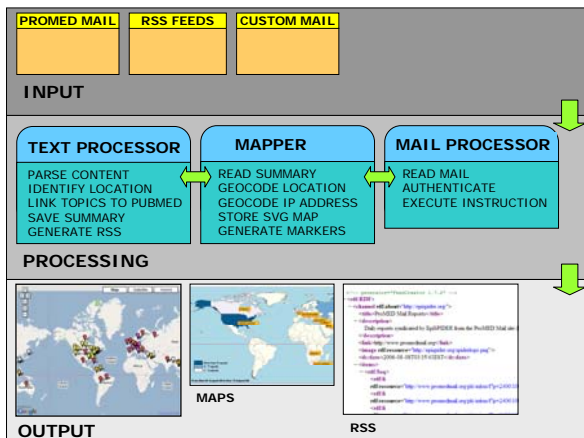


Figure 1. Inputs, processing and outputs of EpiSPIDER. EpiSPIDER processes custom mail instructions through its native mail processor.

infectious disease threats to global health.

Materials and Methods. We use the mailing list of the Program for Monitoring Emerging Diseases (ProMED, <http://www.promedmail.org>) as a source of emerging infectious disease reports. ProMED is a global mailing list with more than 35,000 subscribers and integrates and summarizes reports on outbreaks of diseases of infectious or toxic etiology that affects plants, animals and humans [1]. We also use news sources for health that provide Really Simple Syndication (RSS) feeds like infectious disease sections of the World Health Organization and the European Surveillance Network. EpiSPIDER also obtains RSS feeds containing health