

biocaster : 多言語ニューステキストを利用した 感染症の早期発見および監視

biocaster: Detection and Tracking of Disease Outbreaks from Multilingual News Texts

Nigel Collier and Ai Kawazoe

要旨

SARSやトリインフルエンザのような感染症の発生を早期に発見し、監視・追跡するには、様々な言語で書かれたWeb上のローカルニュースを、各国の政府が責任を持ってモニターする必要がある。BioCasterプロジェクトでは、最新のテキストマイニング技術を活用して多言語のニュース記事をフィルタリングし、構造化された形式で現地語に翻訳するWebポータルを開発する。特に、(1) 多言語知識リソース (オントロジー)、(2) 高性能クラスタコンピュータおよびストレージシステム、(3) 感染症に関するニュース記事と、研究文献や遺伝子データベースにある最新の研究成果をナビゲートする、知的なリンケージシステム等の構築に焦点を当てる。

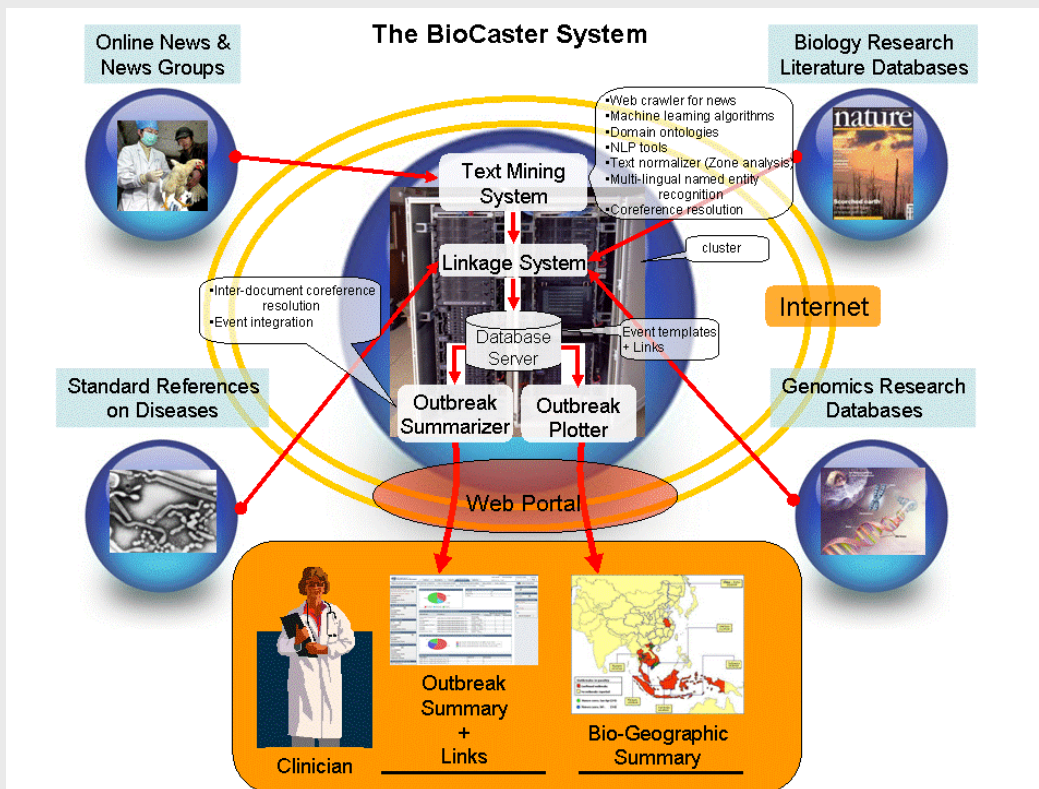
Summary

Early detection and tracking of a possible disease outbreak such as SARS or Avian influenza is a responsibility for governments who are faced with monitoring massive quantities of local news on the WWW in several languages. In BioCaster we are developing a web-portal using the latest text mining technology that can filter news reports in various regional languages and present a summarized translation in the local language. Research will focus on creating: (1) a multi-lingual knowledge resource (ontology), (2) a high-performance cluster computer, (3) an intelligent linkage system for navigating between news about diseases and the latest research findings in the literature and genetics databases.

Purpose

- Enable timely access to disease outbreak news to raise government and public health expert's awareness
- Access to multi-lingual news reports on the Internet using text mining technology
- Integration of bio-geographic information to aid public health workers in analysis of disease spread
- Integration of research literature on disease findings to aid public health workers in deciding on containment and treatment strategy

System Features



biocaster : 多言語ニューステキストを利用した 感染症の早期発見および監視

biocaster: Detection and Tracking of Disease Outbreaks from Multilingual News Texts

Nigel Collier and Ai Kawazoe

日本語ニュースの例

茨城、埼玉両県の養鶏場で昨年流行したH5N2型の鳥インフルエンザで、厚生労働省は11日、感染の疑いのある人が最終的に93人になったと発表した。H5N2型は毒性が弱いうえ、発症者もおらず、2次感染の恐れはないという。(産経Web)

English news

The Ministry of Health in Viet Nam has confirmed a further case of human infection with H5N1. The avian influenza case occurred in a 35-year-old man from Hanoi who was hospitalized with respiratory symptoms on 26 October and died on 29 October. (WHO)

Semantic Markup in News Articles

■BioCaster system recognizes and semantically marks up important entities mentioned in the news text, which are related to disease outbreak event. Examples of such entity classes:

TIME (時間):	<TIME>29 July 2005</TIME>	<TIME>4月6日夜</TIME>	
LOCATION (場所):	<LOC>Viet Nam</LOC>	<LOC>名古屋市</LOC>	<LOC>スマトラ島</LOC>
ORGANIZATION (組織):	<ORG>World Health Organization</ORG>	<ORG>世界保健機関</ORG>	
PERSON (人物):	<PERSON>more than 70 people</PERSON>	<PERSON>18歳の女性</PERSON>	
DISEASE (疾病):	<DISEASE>SARS</DISEASE>	<DISEASE>A型肝炎</DISEASE>	
SYMPTOMS (症状):	<SYMPTOM>cough</SYMPTOM>	<SYMPTOM>微熱</SYMPTOM>	
VIRUS (ウイルス):	<VIRUS>Ebora virus</VIRUS>	<VIRUS>ノロウイルス</VIRUS>	
ANATOMY (身体部分):	<ANATOMY>liver</ANATOMY>	<ANATOMY>腎臓</ANATOMY>	

News Search Interface

The screenshot shows the BioCaster News Search Interface in a Microsoft Internet Explorer browser window. The page title is "biocaster - Communicable disease surveillance from internet news". The interface includes a search bar with the text "News search: (bhe)", a "Search" button, and a "Reset" button. Below the search bar, there are several sections for filtering results:

- 1 Only news articles with:** A list of checkboxes for categories: Case, Named person, Organization, Location, Time, Disease, Symptom, Transmission, Chemical, Drug, Blood product, DNA, Protein, RNA, Anatomy, Virus, Bacteria, and Control.
- 2 Enter specific terms (optional):** A grid of input fields for each category listed in section 1.
- 3 Choose the search operator (optional):** Radio buttons for "Boolean AND (all categories/terms)" and "Boolean OR (any categories/terms)".
- 4 Choose the search domain (optional):** Radio buttons for "All newswires", "General newswires only", and "Public health newswires only".
- 5 Choose date range:** Input fields for "From (YYYYMMDD)" and "To".
- 6 Search** and **Reset** buttons.

■We are developing news search interface which provides users advanced search functionalities to retrieve internet news articles related to communicable diseases and related information.

■Users can search for disease outbreak reports by specifying categories of information and additional search options, from the filtered news collections taken from online news feed every few hours.