

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, Ai Kawazoe, Son Doan, Mike Conway, Reiko Matsuda Goodwin, Dinh Dien, Koichi Takeuchi, Asanee Kawtrakul, Yoshio Tateno, Mika Shigematsu, Kiyosu Taniguchi

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。

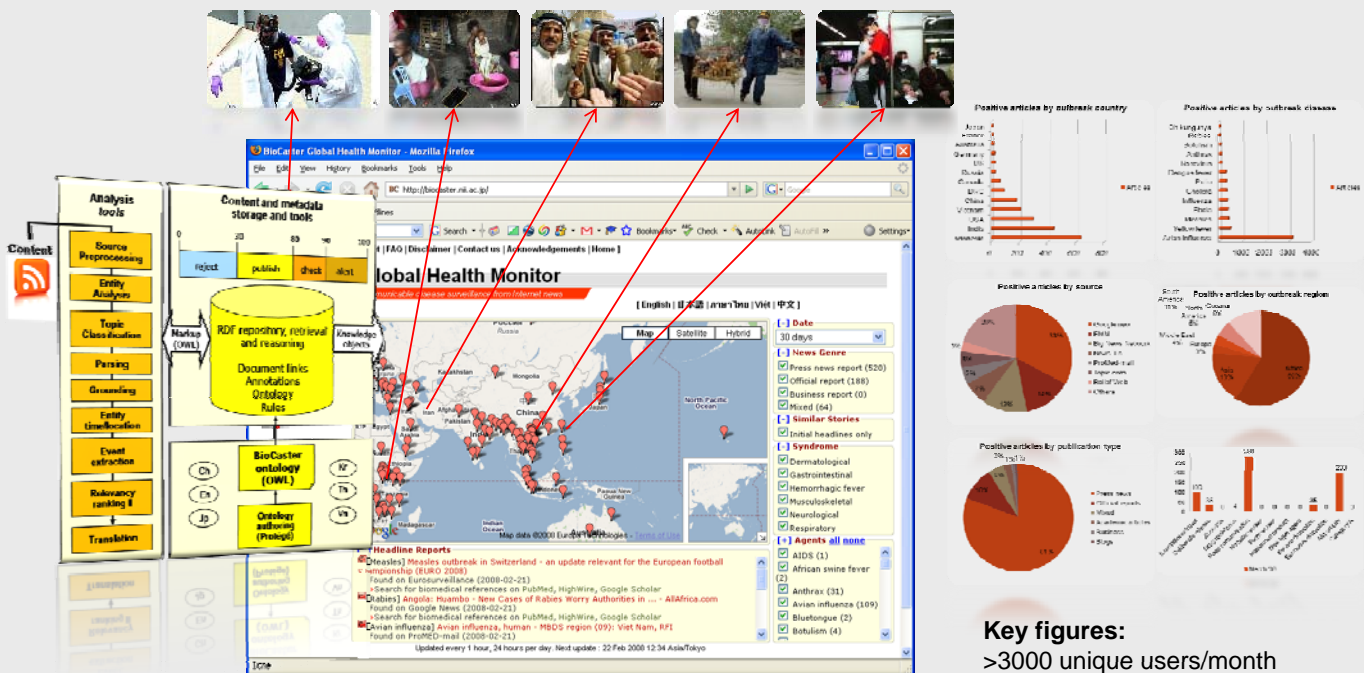
Summary

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 is focusing on creating: (1) a multi-lingual knowledge resource (ontology), (2) a high-performance text mining system, (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 in the analysis of disease spread
- Automatic email alerts to registered users for news items on key topics of interest
- Linkage of rich information sources to help users decide on the significance of the outbreak

System Features



Key figures:

- >3000 unique users/month
- >1700 news sources
- 3000 news reports analysed/day

Find out more at: <http://biocaster.nii.ac.jp>



連絡先: Nigel COLLIER / 国立情報学研究所 Principals of Informatics, Associate Professor
TEL : 03-4212-2536 FAX : 03-3556-1916 Email : collier@nii.ac.jp

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, Ai Kawazoe, Son Doan, Mike Conway, Reiko Matsuda Goodwin, Dinh Dien,
Koichi Takeuchi, Asanee Kawtrakul, Yoshio Tateno, Mika Shigematsu, Kiyosu Taniguchi

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 型は毒性が弱いので、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

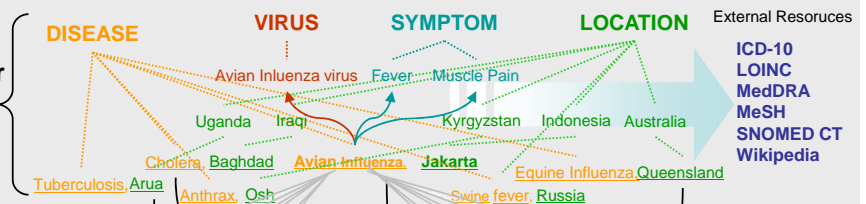
■BioCaster system recognizes and semantically marks up important entities mentioned in the news text, which are related to disease outbreak event. Examples of such entity classes:

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>1 8 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>

Development of a Public Health Ontology

■At the heart of the system is a multi-lingual ontology, which has knowledge for disease surveillance, and serves as a basis for advanced searches and intelligent inferences.

BioCaster
Ontology



BioCaster Ontology ver.2.0

4740 concepts

- Disease: 102
- Syndrome: 7
- Symptom: 120
- Animal: 74
- MicroOrganism: 105
- + Location: 4316

5129 terms

- English
- Japanese
- Thai
- Vietnamese
- Korean
- Chinese
- French
- Spanish

Available for Download at:
<http://biocaster.nii.ac.jp>

