BIOCASTER 公衆衛生保護のためのグローバル・メディア・モニタリング

BioCaster: Global Media Monitoring for Public Health Protection

Nigel COLLIER, Son DOAN, Reiko MATSUDA GOODWIN
Mike CONWAY, Dinh DIEN, Koichi TAKEUCHI, Asanee KAWTRAKUL

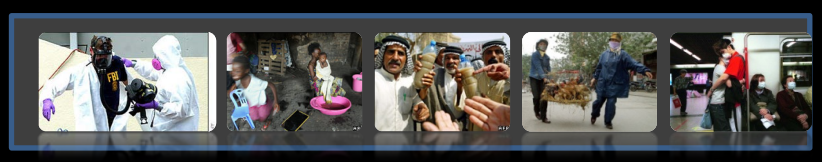
どんな研究？

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応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。

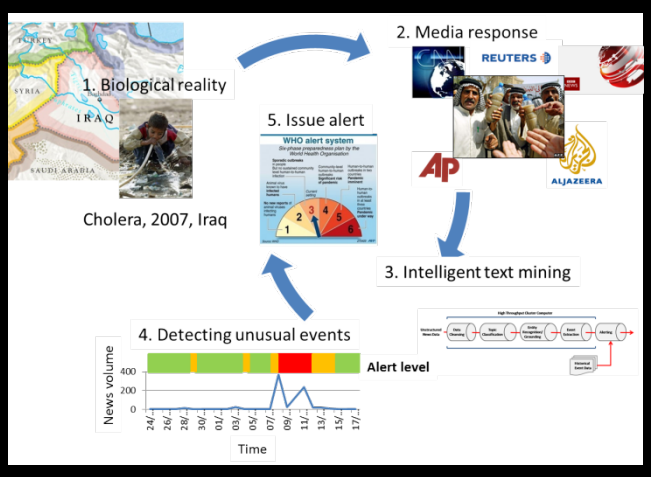
What kind of research?

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.

グローバル・メディア・モニタリングとは？



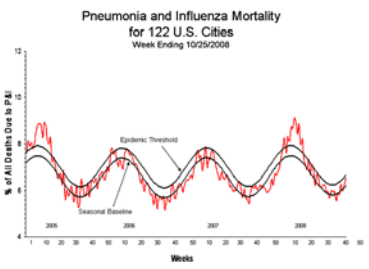
グローバル・メディア・モニタリングは、ニュースなどWebのオープンソースやTwitterのようなソーシャルメディアサイトから、病気の発生や化学物質の流出など公衆衛生に有害なものを素早く察知することを目的としている。BioCasterは、人間や動植物の健康への脅威の発見に活躍している。



課題は？

異常を発見すること…

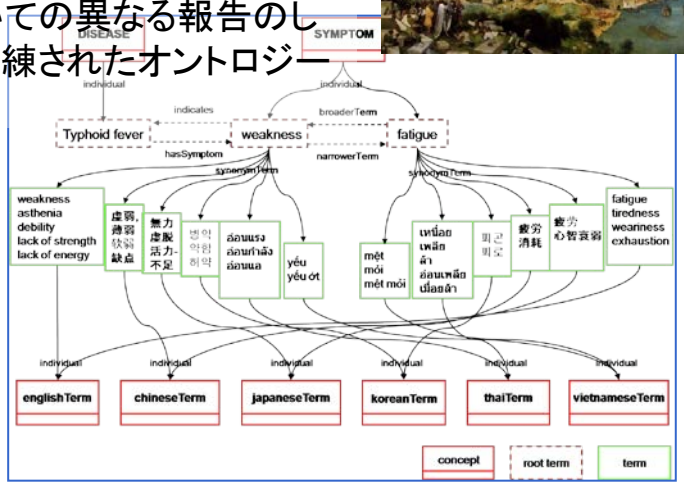
ウガンダでエボラ出血熱？ダブリンでサルモネラ症？ コンピューターはどうやって、ある事象が異常だと認識するのか。本研究では、事象数の統計的アラート用の様々な時系列解析アルゴリズムを分析し、評価する。



CDC data showing epidemic alerting thresholds for influenza

曖昧性を理解すること…

同じ健康状態であっても、たとえば、**インフルエンザ、流感、H5N1、トリインフルエンザ**のように、報告のしかたは多様である。多言語での報告は可能性を表す反面、課題も増やす。研究成果の主なものは、健康状態についての異なる報告のしかたを統一するために、洗練されたオントロジーを作成することである。



Webの限界を理解すること…

ニュースソースの信頼性は様々であり、世界各地に異なる報告パターンがある。事象報告の適用範囲と適時性において、Webには限界があることに気づきはじめた。

情報源を徹底的に調査すること…

ニュース報道、ブログ、検索クエリ…。報告率や集団特性とともに、時間的・空間的精度の異なる可能性のある、メディアを横断するシグナルをどのように統合するのか。どのようにそれらを基準と比較して確認するのか。この疑問については、Grand Challengeが助成するDIZIEプロジェクトの研究で詳しく調査しはじめたところである。

dizie

Biocaster

詳しくは http//born.nii.ac.jpを参照

主要参考文献

1. Collier, N. et al. (2008) “BioCaster: detecting public health rumors with a Web-based text mining system”, BioInformatics, 24(24): 2940-2941, Oxford University Press.

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

BIOCASTER 公衆衛生保護のためのグローバル・メディア・モニタリング

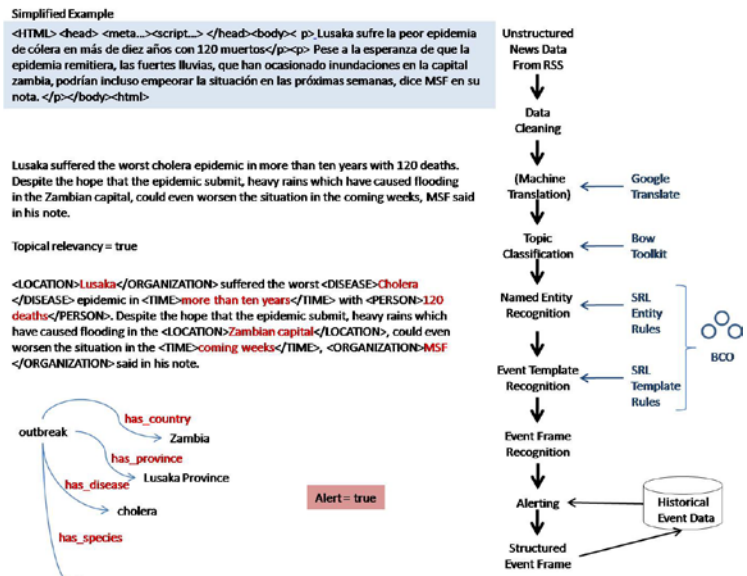
BioCaster: Global Media Monitoring for Public Health Protection

Nigel COLLIER, Son DOAN, Reiko MATSUDA GOODWIN
Mike CONWAY, Dinh DIEN, Koichi TAKEUCHI, Asanee KAWTRAKUL

使用するコア技術は何か？

最適化された特徴選択を使った大容量のデータセットに関して、様々な知的文書処理用先進アルゴリズムを詳しく調べている。主なタスクとしては、テキスト分類、用語認識、事象抽出、事象アラート、可視化がある。システム全体の基礎となっているのは、多言語オントロジー（BioCasterオントロジー、またはBCO）である。BCOは自由にダウンロードして利用することでき、また感染症に関して、体系化された豊富な用語が多数の言語で収録されている。

BioCaster's Semantic analysis pipeline



アウトプットはどのようなものか？

どの程度正確に発生を検出できるのか？

ニュース事象発見の正確さをはかるため、検出率と適時性をProMED-mailと呼ばれる専門家のネットワークと比較した。その結果、BioCasterは0.53の精度で専門家ネットワークによる検出の1週前までに事象を見つけ出せることがわかった。多言語ニュースを含めると0.60であった。

ソーシャルメディアデータの正確さをはかるため、インフルエンザの症状に関するツイッターレポートを調べ、アメリカの研究室基準と比較して相関関係を評価した。相関関係は非常に高く、ほぼ0.99であった。

これらの実験から学んだ教訓は、現在BioCasterとDIZIEのWeb上でのデモに取り入れられている。

一緒に研究しているのは誰か？

パートナーシップは、安全衛生の向上や、研究結果が正確で役立つことを確認するために重要なものである。世界保健機関、欧州疾病対策センター、米国疾病管理センター、日本厚生労働省、英国健康保護局、欧州委員会保健・消費者保護総局、カナダ公衆衛生局をはじめとする多くの国際的公衆衛生機関とともに研究をすすめている。技術パートナーは、タイのカセサート大学、ヴェトナム国立大学、岡山大学などである。

BioCasterは、JSTさきがけ基金から、DIZIEは、国立情報学研究所のGrand Challengeプロジェクト基金から助成金を受けている。

Biocaster

詳しくは http//born.nii.ac.jpを参照

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