

## 感染症への適切な対応の基盤となる多言語オントロジーの開発

研究代表者: Nigel Collier

### 1. 共同研究者

[国立情報学研究所]	藤山 秋佐夫、川添 愛、Reiko Matsuda Goodwin
[国立感染症研究所]	谷口 清洲、重松 美加
[岡山大学]	竹内 孔一
[国立遺伝学研究所]	舘野 義男
[カセサート大学 (タイ) ]	Asanee Kawtrakul
[ベトナム国立大学 (ベトナム) ]	Dinh Dien
[西オーストラリア大学 (オーストラリア) ]	Roberto Barrero

### 2. 研究目標

感染症の流行は世界的な懸案事項であるが、アジア太平洋地域には監視システムのためのインフラストラクチャが無い。このことが、近年の H5N1 型トリインフルエンザ等の感染症を分析し、コントロールする上で障害となっている。BioCaster プロジェクトは、この地域の監視体制改善のために、テキストマイニングを基盤とする自動的な監視システムを開発することを目的とする。このシステムは、アジアの諸地域の言語を使用して、インターネット上のニュースやその他のオンラインの情報源を自動的にモニターするものである。システムの中核をなすのはアプリケーション・オントロジーであり、これはテキストマイニングにより抽出された事実に対する（多言語間の）高度な検索を可能にすると同時に、イベントの重要度・優先度を評価するための知的な推論を行うという、二つの目的を持つ。BioCaster オントロジーは病原体の優先度、病原体が起こす疾病、および環境から宿主の遺伝子を介して人体へと至る感染ルートに焦点を置いた多言語オントロジーである。

### 3. 平成19年度までの研究進捗及び主要成果物

#### 【研究進捗】

平成 18 年度は、BioCaster オントロジーの第一版を公開した。このオントロジーでは、言語処理技術に基づく生物医学分野の知的情報システムをサポートする目的で、27 の病原体に関する完全な記述を 6 カ国語で行っている。このオントロジーは BioCaster Web ポータル上で公開している。Web ポータルには一ヶ月あたり 1500 人以上のユーザが訪れており、現在増加傾向にある。

また平成 19 年度には、病原体の数を 90 に増やし、6 カ国語（英語、日本語、タイ語、ベトナム語、韓国語、中国語）の用語のカバレッジも増やした。また、Web ポータル上でダウンロード可能なバージョンを公開した。この研究の成果は学術誌や国際会議で発表している。

#### 【主要成果物】

<平成 18 年度>

多言語オントロジー第一版

- 27 の重要伝染病の病原体、症状、感染経路、宿主生物について記述
- 6 つのアジア太平洋地域の言語（中国語（簡体字）、英語、日本語、韓国語、タイ語、ベトナム語）の用語を登録
- MeSH、LOINC、SNOMED CT などのオンラインの専門オントロジーや、Wikipedia のような一般向け参照用リソースへのリンクを付加
- Web ポータル (<http://biocaster.nii.ac.jp>) にてオントロジーのオンライン検索機能を提供

<平成 19 年度>

- 記述対象の病原体を 90 に増加
- 6 カ国語（英語、日本語、タイ語、ベトナム語、韓国語、中国語）の用語のカバレッジを 40%ほど増加
- ウェブオントロジー言語（OWL）で書かれたダウンロード用オントロジーを公開
- 一ヶ月あたり 1500 人以上のユーザが訪問（現在）

#### 4. 平成20年度以降の展開

平成 20 年度以降は、1)オントロジーの拡張 2) オントロジーと現在開発中の感染症監視システムとの統合 を行う予定である。

##### 1) オントロジーの拡張

現行のオントロジーを、言語と概念の両面で拡張する。まず、言語の面では、新たにインドネシア語とマレー語を対象言語に加え、これらの言語で使用されている感染症関連の用語をオントロジーに含める。インドネシア周辺は、現在 H5N1 型インフルエンザの流行が問題になっている地域であり、これらの地域で使用されている言語の知識をオントロジーに組み込むことは、感染症監視システムがローカルな情報を迅速に取得する上で重要である。

概念の面では、記述対象の病原体の数を 150 に拡張し、また症候群や症状に関する部分の拡張も行う。後者に関しては、米国ピッツバーグ大学の Wendy Chapman 教授との共同研究を計画している。現在、症候群や症状の分類方法は多様に存在し、目的に合わせて適切な分類を選択するのは容易ではない。Chapman 教授はカルテなどの医学報告書に現れる症状の記述の分析に携わっており、本研究が対象にしているニュース記事上の症状の記述の分析と共通するところが多いため、共同研究により問題とその解決についての深い理解が得られると期待している。

##### 2) 感染症監視システムとの統合

現在当グループでは、BioCaster という、Web 上のニュース記事を利用した感染症監視システムの開発を行っている。このシステムは、テキストマイニング技術を基盤とし、大量のニュース記事から感染症の発生イベントを報じているものを見つけだし、情報を構造化されたデータとしてユーザに提供することを目的とするものである。20 年度はこのシステムに、感染症に関する

る知識および言語の知識の基盤として多言語オントロジーを組み込む予定である。多言語オントロジーと感染症監視システムの統合により、世界各国の公衆衛生分野の専門家、および政策決定者に対し、感染症の最新情報への母国語によるアクセスを提供するシステムが構築でき、また言語間で一対一に対応しない用語間の関係にも対処することができる。また、これは現在広く使用されている単純なキーワード検索システムから、次世代の意味（セマンティックス）による検索システムへの移行においても重要となる。

## 5. 研究経費

平成17年度実績： 0千円  
平成18年度実績： 7000千円  
平成19年度見込： 7000千円

## 6. 平成19年度の研究成果

### (1) 知見・成果物・知的財産権等

BioCaster オントロジー (<http://biocaster.nii.ac.jp/index.php?page=downloads&submit>)

### (2) 成果発表等

#### <論文発表>

##### 【学術論文】

1. Kawazoe, A., Chanlekha, H., Shigematsu, M. and Collier, N. (2008), "Structuring an event ontology for disease outbreak detection", in BMC Bioinformatics (in press).
2. Collier, N., Kawazoe, A., Jin, L., Shigematsu, M., Dien, D. Barrero, R., Takeuchi, K. and Kawtrakul, A. (2007), "A multilingual ontology for infectious disease surveillance: rationale, design and challenges", Language Resources and Evaluation, Elsevier, DOI: 10.1007/s10579-007-9019-7.
3. Collier, N., Kawazoe, A., Son, D., Shigematsu, M., Taniguchi, K., Jin, L., McCrae, J., Chanlekha, H., Dien, D., Hung, Q., Nam, V., Takeuchi, K. and Kawtrakul, A. (2007), "Detecting Web Rumours with a Multilingual Ontology-Supported Text Classification System", Advances in Disease Surveillance, Vol. 4, pp. 242.
4. Kawazoe, A., Jin, L., Shigematsu, M., Barerro, R., Taniguchi, K. and Collier, N. (2006), "The development of a schema for the annotation of terms in the BioCaster disease detection/tracking system", Olivier Bodenreider (ed)., Proceedings of the International Workshop on Biomedical Ontology in Action (KR-MED 2006), Baltimore, Maryland, USA, November 8, pp. 77-85,
5. Collier, N., Kawazoe, A. Shigematsu, M., Taniguchi, K., Jin, L., McCrae, J., Dien, D., Hung, Q., Takeuchi, K., Kawtrakul, A. (2007), "Ontology-driven Influenza

Surveillance from Web Rumours”, in proceedings of the 2007 Options for the Control of Influenza VI (Options), Toronto, Ontario, Canada, June.