

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

1. 研究代表者名

[国立情報学研究所 情報学プリンシプル研究系] Nigel Collier

2. 共同研究者

[国立情報学研究所] 藤山 秋佐夫 舘野 義男

[国立感染症研究所] 谷口 清洲 重松 美加

[岡山大学] 竹内 孔一

[国立情報学研究所] 川添 愛 金 麗華

[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

3. 平成 18 年度の研究実績報告

3-1. 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。

3-2. 平成 18 年度研究成果（概要）

18 年度はオントロジー構築の第一段階を完了した。27 の重要伝染病に焦点を置き、これらの病原体、症状、感染経路、宿主生物について完全な記述を行った。この記述には各概念の定義に加え、概念間の関係も含まれる。また、このオントロジーの多言語用リソースとしての側面を確立するため、今年度は 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）において、疾病、病原体、症状のカテゴリに入る各概念を表す用語を調査・収集してオントロジーに含めた。収集された用語は各言語のネイティブ・スピーカー（疫学の専門家含む）によってチェック済みである。更に、オントロジー内の各概念に対して、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などへのリンクも付加した。

構築したオントロジーは現在 Web ポータルとして一般に公開しており、ユーザはオンライン検索などの機能を使用することができる。まもなく、ウェブオントロジー言語 (OWL) で書かれたダウンロード可能なバージョンを公開する予定である。

3-3. 今後の展開

今年度以降は、次のようにオントロジーを拡張していく計画である。

- オントロジーがカバーする病原体の種類を、1年に40ずつ追加する。人間に及ぼす病気、宿主、症状などの情報も合わせて記述する。各々の概念に対する用語を、現行の6ヶ国語で記述する。
- オントロジーを拡張し、より多くの言語に対応できるようにする。特に、繁体字中国語、インドネシア語への拡張を優先的に行う。
- 病原体の人体への感染にかかわる宿主遺伝子についての情報を追加する。
- 病原体が人体に感染する際にターゲットとなる器官や組織に関する情報を追加する。

また、一般公開可能なオントロジーを、1年に2回ずつ更新し、研究者コミュニティ向けにリリースする。更に、オントロジーの詳細についての報告を、学術誌などで発表する予定である。

成果発表実績

<知見・成果物・知的財産権等>

1. Nigel Collier, Ai Kawazoe, Lihua Jin, Mika Shigematsu, Dinh Dien, Roberto Barrero, Koichi Takeuchi, Asanee Kawtrakul, The BioCaster Ontology database, available at <http://biocaster.nii.ac.jp>.

[論文発表]

<学術論文>

1. Nigel Collier, Ai Kawazoe, Lihua Jin, Mika Shigematsu, Dinh Dien, Roberto Barrero, Koichi Takeuchi, Asanee Kawtrakul, "A multilingual ontology for infectious disease outbreak surveillance: rationale, design and challenges", to appear in the Journal of Language Resources and Evaluation, Elsevier, 2007.

<会議録>

1. Nigel Collier, Ai Kawazoe, Mika Shigematsu, Kiyosu Taniguchi, Lihua Jin, John McCrae, Dinh Dien, Quoc Hung, Koichi Takeuchi, Asanee Kawtrakul, "Ontology-driven influenza surveillance from Web rumors" to appear in Proceedings of the 2007 Options for the Control of Influenza VI (Options), Toronto, Ontario, Canada, June 17th-23rd, 2007.