

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 24 日現在

機関番号：82636

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23500199

研究課題名(和文)サイバー・フィジカルシステムのためのスケーラブルなセマンティックサービス基盤技術

研究課題名(英文)Scalable Semantic Service Infrastructure for Cyber-Physical System

研究代表者

岩爪 道昭(Iwazume, Michiaki)

独立行政法人情報通信研究機構・ユニバーサルコミュニケーション研究所・統括

研究者番号：80319756

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：近年急速に流通しているOpen Linked Dataのためのサービス基盤技術の研究開発に取り組んだ。(1)イベントデータのモデリング手法を開発し、これをDBpediaなどの意味的資源に適用し、プロパティ情報に着目した定式化を行った。(2)(1)で定式化した意味的プロパティ情報のランキング手法を定式化・実装し、DBpediaを対象にした実証実験を行い、その有効性について確認した。(3)Open Linked Data等の意味的資源を収集・蓄積・解析するためのクローラー、分散データストア等のミドルウェアおよびそれを支える大型計算機システムの動態情報を可視化を試作し、動作検証を行った。

研究成果の概要(英文)：In recent years, a number of Semantic Services have been actively open to public as common resources on the Web by the effort of Linked Open Data community project. Due to this, we need a good method to search necessary data from those services, depending on various purposes. In this study, we propose three methods:(1)to formalize Linked Data based-on event ontology, (2) to model rank the results retrieved by a SPARQL query (especially, a SELECT query),using the information about the frequency of each property in a data set and the links between RDF resources and (3)to monitor a large-scale semantic information infrastructure based on a PC cluster systems.

研究分野：知能情報処理

キーワード：Open Liked Data Semantic Web ビッグデータ 大規模分散データストア

1. 研究開始当初の背景

近年、センサ技術およびネットワーク技術の進展・普近年、ライフログ、ソーシャルネットワークサービスや Internet of things と呼ばれるセンサ、電子タグ等と接続したネットワークから非定型かつ大量のデータがリアルタイムに生成されるようになってきた。このようなデータは、短時間のうちに数億から数百億のデータエントリ数、ペタバイト級のデータサイズに及ぶことから「ビッグデータ (big data) 」と呼ばれる。ビッグデータの蓄積、解析し、その結果に基づき実世界を最適化・制御するサイバー・フィジカルシステム (以下 CPS) が新しいイノベーション創出の基盤として関心が高まっている。

一方、最近、セマンティック Web の研究分野では、生物、医学、地球環境計測および電子政府等の分野において、個別に蓄積・管理されていた各種データが、セマンティック Web の仕様に準拠した RDF と呼ばれる主語-述語-目的語の 3 つ組み形式 (トリプル) で Web 上に公開され、他の公開 RDF データ、オントロジーと相互リンクされることで、大規模意味資源が形成されつつある。このような意味資源は、Linking Open Data (以下 LOD) と呼ばれている。この大規模知識資源を背景として、セマンティック Web 研究の加速と多様な社会サービスの創出が期待されている。

我々は、CPS の汎用的な意味処理基盤として要求されているスケーラビリティ、柔軟性、モジュール性、オープン性の観点から、上述の LOD をはじめとするセマンティック Web 技術が、重要な役割を果たすことになると考えて本研究課題に取り組むこととした。

2. 研究の目的

上記の背景を踏まえ、我々はこれまで研究成果を発展・融合させ、次の課題に取り組むこととした。

(1) 超大規模意味資源のサービス指向モデリング技術: CPS を構成する異なる粒度、時間スケールを持つ超大規模意味資源に対し、システム設計・管理者および利用者が、明示的に意味のある単位で挙動や機能を切り出し、サービスとして設計・実装可能とする分析モデリング技術。

(2) スケーラブルなサービス疎結合技術: (1) によってモデル化された多様なサービス群を、意味資源の性質・機能および利用者やシステム全体の要求仕様 (ポリシー) に応じ、柔軟に合成、連携、協調、再構成するためのトランススケールなサービス疎結合技術。

(3) トランススケールサービスモニタリング技術: (2) によって有機的に結合された複雑かつ超大規模な CPS サービス空間における、個々のサービスや CPS の系全体の挙動・機能、リネージ (由来となるデータや下位サービス) 等を、利用者やシステム設計者・管理

者が理解可能な形で可視化及び挙動シミュレーションし、提供サービスの質を評価可能とするサービスエンビジョニング技術。

3. 研究の方法

(1) 大規模意味資源のサービス指向モデリング: モデル化の対象となる多様なデータ収集・分析を行った。ヘルスケア分野 (心電、脈拍、呼吸数、経皮的動脈血中酸素飽和度等)、地球環境リモートセンシング分野 (降雨レーダ情報)、エネルギー分野 (データセンター設備のエネルギーマネジメント)、生活安全分野 (降雨レーダ) の各種データをサンプルとして収集し、イベントオントロジーにおける a) 状態変化、b) 時間上の存在変化、c) 空間上の存在変化、d) 基数変化、e) 比較と同一性変化の意味機能の観点から微細データに意味付与する手法検討した。

(2) スケーラブルなサービス疎結合技術:

多様な粒度のサービスを緩やかに連携・結合させるため、意味的な Link 情報を介したサービス間の関係性の発見および連携手法の開発に取り組んだ。具体的には、Open Liked Data を知識源とした様々な情報サービスの連携を、意味的なリンクの重要度に応じて実現するため、そのランキング手法の定式化と実証検証を行った。

(3) トランススケールなサービスモニタリング技術: Linked Open

大量の意味的データ蓄積、解析とそれに基づくサービスの運用においては、それを支える情報基盤システム自体が様々なログデータを大量に生成する。障害時発生時のトラブルの切り分けや計算処理の高速化のためのシステムチューニング等において、これらのログ情報の有効なフィードバックを得ることが重要になってくる。そこで、計算機、ネットワーク環境、データベース等基盤ソフトウェアの動態情報を一体的に集約、管理するため次の機能の実現に焦点を当てた。

(a) データ集約: 数 100 ~ 1000 ノード規模のサーバおよびその上で動作する基盤ソフトウェアから出力されるログ情報を毎秒 ~ 日単位で収集する機能。

(b) データ管理: 刻一刻と生成される多様かつ膨大なログデータをデータベースに格納し、必要に応じてデータ単位を要約し管理する機能。

(c) データ可視化: 収集されたを動態情報として分かりやすく可視化する機能。

4. 研究成果

スケーラブルなセマンティックサービス基盤技術の実現に向けて、平成 23 年度は、現在急速に増加、流通しつつある Linked Data に焦点を当て研究に取り組んだ。平成 24 年度は、近年急速に注目されるようになったビッグデータに対応する大規模な計算機基盤、ミドルウェアソフトウェア群のサービス化についてより具体的なアプリケーション

ン、データを対象に実証検証を行った。平成 25 年度は、さらにビッグデータの中でも Open Linked Data とそれを支えるビッグデータ基盤技術の実証検証に焦点を当て研究開発に取り組んだ。最終年度の平成 26 年度は、より大規模な計算機環境を用いた総合実証を行った。以下では、各課題の成果について詳述する。

(1) 大規模意味資源のサービス指向モデリング：現実世界における事象や行為の発生や状態の変化を扱うためにイベントデータの形式的モデリング手法について基礎検討を行った。特に、イベントが含む多様な情報を記述してデータ化することを目的に、イベントの構成物、意味機能、およびイベント間の関係による 3 つの意味的視点から新しいイベントタイプの分類を行い、ソート階層(概念階層)を用いたイベントオントロジーの定式化を行った。センサ機器による実世界情報をモデリングし、Open Linked Data として相互運用するための方式を具体的に検証するため、特にヘルスケア/ライフログ分野に焦点を当て、GPS ナビゲーション機器、加速度センサ、歩数計、心拍計等の各種機器を用いた検証データの収集に着手した。平成 25 年度は、平成 23~24 年度に検討したイベントデータの形式的モデリング手法を、DBPedia などの意味的資源に適用し、プロパティ情報に着目したランキング手法の定式化を行った。またこれと並行して、応用アプリケーションとして、平成 24 年度に収集したライフログデータを対象にして、当該モデリング手法の適用可能性についても確認した。

(2) スケーラブルなサービス疎結合技術：意味的な Link 情報を介したサービス間の関係性の発見および連携手法を実現するため、まずは意味的データの収集・蓄積手法に着手した。具体的には、Java バージョンマシン上でスケーラブルに動作する分散仮想クローラーとそれらを蓄積・検索可能とする分散データストアを 4 ノード、8 コアからなる小規模な PC クラスタ上で試作した。また、DBPedia のデータセットページをシードとして 17,163 トリプルを収集し、これを元に 5 億トリプルの疑似データを生成し、分散データストアに適用することで実効的な時間でスケールできることを確認した。さらに、収集した Linked Data に対して HITS アルゴリズムによるリンク解析を行うことで、意味的情報を拡張する手法を開発し、意味的サービス間の連携への適用可能性を確認した。平成 24 年度は、前年度に取り組んだ Open Linked Data に対するリンク解析の高度化・大規模化に取り組んだ。具体的には DBPedia の全データを対象としてプロパティ情報意味特徴を踏まえたランキングアルゴリズムの定式化した。また 100 億リンク規模の Web グラフを PC クラスタ上で高速解析する手法についても検討を行った。意味的なリンク解析のさらなる高度化・大規模化に取り組んだ。

具体的には、高度化に関しては、(1)で定式化した意味的プロパティ情報のランキング手法を実装し、DBPedia 全データを対象にした実証実験を行い、従来手法と比較し、その有効性について確認した。

(3) トランススケールなサービスモニタリング技術：Open Linked Data を始めとする Web 情報を収集するための計算機基盤、クローラー、分散データストア等のミドルウェアが生成するログデータや動態情報を集約的に可視化するプロビジョニング環境を試作した。特にクローリングサービスにおいては、可視化結果に基づくクローラーの各種パラメータのチューニングの結果、従来比で収集性能 2 倍に向上し、当該手法の有効性を確認した。平成 25 年度は、平成 24 年度に試作した意味的資源のクローリング情報の可視化システムを改良、拡張子、オープンリンクドデータや HTML 形式の文書を収集、蓄積、解析するための大規模クローラー、分散データストア等のミドルウェアおよびそれがどうする大型計算機のハードウェアの動態情報を可視化し、安定的に運用を支援するためのモニタリングシステムを試作し、動作検証を行った。

(4) 総合検証：サイバー・フィジカルシステムにおいては、当研究課題が対象にしてきた Open Linked Data 等のネット上の情報だけでなく、センサーネットワーク等からの実世界情報の収集・蓄積と活用のためのサービス化が重要である。このようなデータは生成速度が速く、構造化されていないため、従来のバッチ処理型のアプローチでは対応が困難であった。そこで、このような大規模なストリームデータをメモリ上でリアルタイムまたは準リアルタイムに解析し、再利用可能な形に構造化、永続化するデータベース技術が不可欠である。またそれらを実際に稼働させるためには計算機基盤も従来のハードディスク主体のストレージ装置では不十分であり、メモリや SSD 等のテイレンテンシーなストレージを搭載した計算機が不可欠である。このような状況を踏まえ、研究代表者所属機関で導入した大規模共有メモリを含む大型計算機設備を用い総合実証を行った。また、平成 25 年度末までに取り組んできたインターネット上の Web 情報、Open Linked Data に加え、降雨レーダのデータを想定したインメモリデータベースの実装検証を行いその有効性を確認した。

5. 主な発表論文等

〔雑誌論文〕(計 6 件)

兼岩 憲, 岩爪 道昭, セマンティック Web のためのイベントオントロジー, コンピュータソフトウェア(日本ソフトウェア科学会論文誌), 査読有, Vol.28, No.3, 2011, 153-166
Kanakano Onishi, Ichiro Kobayashi,

Information Enhancement on a Focused Object Using Linked Data, Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有, Vol. 16, No.1, 2012, 4-12

Ken Kaneiwa, Riichiro Mizoguchi, A Time and Situation Dependent Semantics for Ontological Property Classification, IEICE Transactions on Information and Systems, 査読有, Vol. E94-D, No. 3, 2011, 639 - 647

藤井 秀明, 原口 弘志, 泥谷 誠, 岩瀬 高博, 岩爪 道昭, ビッグデータ時代における情報の巨大集積化・並列分散処理に関する研究開発動向, 日本ソフトウェア科学会論文誌「コンピュータソフトウェア」, 査読有, Vol.29, No.1, 2013, 130 - 151

岩瀬 高博, 藤井 秀明, 田中 康司, 原口 弘志, 泥谷 誠, 岩爪 道昭, 大規模情報基盤のための分散キー・バリューストア okuyama による分散共有ストレージの構築, 人工知能学会誌, 査読有, 2013, Vol.28.No.6, 104-113

Michiaki Iwazume, Ken Kaneiwa, and Ichiro Kobayashi, KC3 Browser: Semantic Service Mash-up for Global Knowledge Sharing and Discovery, In Proceedings of the 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD 2012), Springer, 査読有, Vol.443, 2013

〔学会発表〕(計 16 件)

岩爪 道昭, 河原 大輔, 兼岩 憲, 赤峯 亨, 加藤 義清, 大西 可奈子, 小林 一郎, Web サイエンス基盤のための大規模意味資源のスケラブルな集積化の試み, 2011 年度人工知能学会全国大会, 2011 年 6 月 3 日, 岩手県民情報交流センター (アイーナ)

大西 可奈子, 小林 一郎, 岩爪 道昭, Linked Data を利用した対象文章の情報拡張への取り組み, 2011 年度人工知能学会全国大会, 2011 年 6 月 3 日, 岩手県民情報交流センター (アイーナ)

大西 可奈子, 小林 一郎, リソース間の有意な関係を見出す Linked Data 解析手法への取り組み, 第 74 回情報処理学会全国大会 2012 年 3 月 6 日, 名古屋工業大学

岩爪 道昭, アジア情報 HUB プロジェクト (第一報), 2012 年度人工知能学会全国大会, 2012 年 06 月 12 日, 山口県・山口県教育会館

藤井 秀明, 岩瀬 高博, 原口 弘志, 泥谷 誠, 岩爪 道昭, ハイパフォーマンス・エラスティック・クロール

岩瀬 高博, 藤井 秀明, 泥谷 誠, 岩爪 道昭, ビッグデータインメモリ, 2012 年度人工知能学会全国大会, 2012 年 06 月 12 日, 山口県・山口県教育会館

Michiaki Iwazume, Asian Information HUB Project, The 6th IEEE International Workshop on Middleware Architecture in the Internet (Midarch2012), 2012 年 07 月 16 日, トルコ共和国イズミール・スイスホテル会議場 (招待講演)

岩爪 道昭, 岩瀬 博, 田中, 藤井, 原口, 泥谷: ビッグデータの集積化とその活用を支える大規模情報インフラの構築・運用, 2013 年度人工知能学会全国大会, 2013 年 6 月 5 日, 富山県富山市・富山国際会議場

一瀬 詩織, 小林 一郎, 岩爪 道昭, 田中康司: DBpedia における SPARQL 検索結果のランキング手法, 2013 年度人工知能学会全国大会, 2013 年 6 月 5 日, 富山県富山市・富山国際会議場

田中康司, 原口 弘志, 藤井 秀明, 泥谷 誠, 岩爪 道昭, 大規模リンク解析の高速化手法の検討と計算機基盤構築, 2013 年度人工知能学会全国大会, 2013 年 6 月 5 日, 富山県富山市・富山国際会議場

藤井 秀明, 原口 弘志, 田中康司, 泥谷 誠, 岩瀬 高博, 岩爪 道昭, 大規模 Web クローラーの運用事例にみる課題と対策, 2013 年度人工知能学会全国大会, 2013 年 6 月 5 日, 富山県富山市・富山国際会議場

Michiaki Iwazume, Community-driven and Ontology-based Biological Knowledge Management, a Hybrid Approach to Harnessing Collective Intelligence, SNPD2013, 米国ハワイ州・シェラトンワイキキ会議場

Siori Ichinose, Ichiro Kobayashi, Michiaki Iwazume, Koji Tanaka, Ranking the Results of DBpedia Retrieval with SPARQL Query, JIST2013, 2013 年 12 月 13 日~14 日, 韓国ソウル市・COEX コンベンションセンター

一瀬詩織, 小林 一郎, 岩爪 道昭, 田中康司, RDF リンクの構造を考慮した SPARQL クエリ検索結果のランキング手法, 第 3 階 AEG Web インテリジェンスとインタラクション研究会, 神奈川県横浜市・さくら WORKS, 2013 年 12 月 13 日~14 日, 神奈川県横浜市・さくら WORKS

Michiaki Iwazume, Koji Tanaka, Makoto Hijiya, Takahiro Iwase, Hideaki Fujii, Hiroshi Haraguchi, Big Data in Memory: Benchmarking In Memory Database Using the Distributed Key-Value Store for Machine to Machine Communication, 15th IEEE/ACIS International

Conference on Software Engineering,
Artificial Intelligence, Networking
and Parallel/Distributed
Computing(SNPD2014), 2014年6月30
日~7月2日, 米国ネバダ州ラスベガス
市・ハラーズカジノホテル会議場
Michiaki Iwazume, Takahiro Iwase,
Koji Tanaka, Hideaki Fujii ,
Benchmarking In Memory Database
Using the Distributed Key-Value Store
for Constructing a Large Scale
Information Infrastructure, The 8th
IEEE International Workshop on
Middleware Architecture in the
Internet(MIDARC2014), 2014年7月21
日~7月25日, スウェーデン王国ヴェス
テロース市・アロスコンヴェンションセ
ンター

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕

ホームページ等 なし

6. 研究組織

(1) 研究代表者

岩爪 道昭 (IWAZUME, Michiaki)

独立行政法人情報通信研究機構・ユニバー
サルコミュニケーション研究所・統括

研究者番号: 80319756

(2) 研究分担者

兼岩 憲 (KANEIWA, Ken)

電気通信大学・情報理工学(系)研究科・教
授

研究者番号: 00342626

(3) 連携研究者

小林 一郎 (KOBAYASHI, Ichiro)

お茶の水女子大学・大学院人間文化研究
科・教授

研究者番号: 60281440