

機関番号：82626

研究種目：基盤研究（A）

研究期間：2008～2010

課題番号：20240010

研究課題名（和文）不均質なペタバイト級時空間センサデータの統合利用基盤

研究課題名（英文）Integrated Service Infrastructure for Petabyte-Scale Sensor Data with Spatiotemporal Diversities

研究代表者

小島 功 (KOJIMA ISAO)

独立行政法人産業技術総合研究所・情報技術研究部門・研究グループ長

研究者番号：00356982

研究成果の概要（和文）：

本研究では、様々な衛星センサを中心とした時間・空間的に不均質な情報の利用基盤として、異種大規模の情報を効果的に検索・連携することで新たな知見創出に貢献する技術研究開発を行った。インターネット等の膨大な情報と衛星情報の処理結果等、異種の情報を相互に連携させることで事象の正確な理解に役立つ技術を開発すると共に、これら異種膨大な情報に対し OGC 標準に基づく分散検索システムを開発、サービス提供を可能とした。

研究成果の概要（英文）：

This research theme aims to provide the technology for constructing the infrastructure for petabyte-scale sensor data with spatiotemporal diversities. The result includes the method to associate web data and satellite sensors to understand events in real world. Distributed information retrieval method for heterogeneous metadata based on OGC standards is also developed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	11,500,000	3,450,000	14,950,000
2009年度	10,500,000	3,150,000	13,650,000
2010年度	10,500,000	3,150,000	13,650,000
総計	32,500,000	9,750,000	42,250,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：データベース(DBMS)、データ統合

1. 研究開始当初の背景

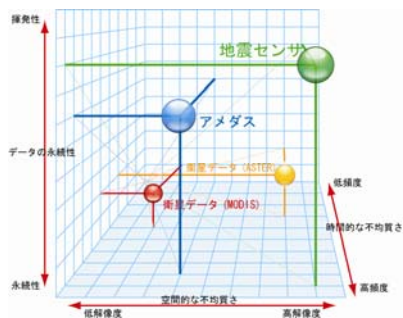
温暖化に代表される環境問題は、地球規模での問題解決が必要であり、全世界的に分散している、さまざまな情報源からの大量の情報を、効果的な形で統合・連携させて問題解決の一助にすることが不可欠である。例えば、人工衛星からの画像やリアルタイムで取得される分散センサからのデータは、時間・空間的な広がりを持つと同時に、多様かつ膨大な量にのぼり、データの効率的な管理や高速な検索、広域での分散処理やデータマイニング等技術課題が多い。なかでもデータ統合、しかも広域に分散した大規模なセンサの統合は、情報の組み合わせによる知見の創出の

ために不可欠な技術として、実用的な基盤の確立が求められており、地球科学応用に限らず、広く計算機を用いた科学研究を促進する基盤(e-Science 基盤)に重要な技術である。

この課題に対し、世界中で様々な要素技術や基盤システムの研究開発が行われている。要素技術としては、データ統合、センサを扱うストリームデータベース、分散データベース処理等が従来から研究が盛んである。現在は、特に地球観測応用に対して、これらの技術を広域の環境で結びつけた統合システムの研究開発や、GEOS, OGC 等データの相互利用のための標準化活動が盛んである。産総研においても、GEO Grid プロジェクトとして、

ASTER と呼ばれる衛星センサ情報基盤とし、多様なデータや計算をグリッド技術で結び付ける研究開発をしている。

さて、GEO Grid で扱う ASTER と MODIS と呼ばれる衛星センサは、画像の解像度や撮影の時間間隔に大きな差があり、利用者が必要とする空間や時間幅を指定したとき、各センサの画像データを管理するデータベースから得られる検索結果のデータ量には大きな差が生じる。例えば画像のつきあわせによる校正を行おうとすると、ある1枚の画像に対して、ある範囲では一枚も画像が見つからないとか、逆にある範囲に膨大な画像が見つかる、という状況が多々生じる。利用者の意図しない膨大な数のデータを扱う必要が生じたり、逆に条件がデータに比して詳細すぎ、本来見つかるべきデータがうまく見つからないなど、データの空間的な解像度やその発生時間間隔、サイズなどが大きくばらつくことに起因する問題が発生し、情報の高度利用を妨げていることが明らかになってきた。形式の似通った衛星画像同士でもこうした問題があるように、沢山のセンサからの地球観測情報や、多様な地物情報などを組み合わせる場合、この問題は大きくなる一方である。



(図1 情報源の種類と不均質さ)

そこで、空間・時間のばらつきが極めて大きい「不均質」なデータ、それも大量の情報を常時発生するようなセンサデータ等の情報源を対象として、情報を効果的に扱え、解析や知見の獲得という利用ができる基盤を提供し、標準化も含めて広く普及させようとするのが、本研究の背景であった。

2. 研究の目的

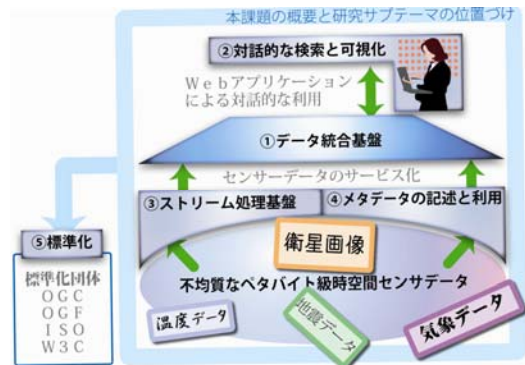
本研究では、時間・空間軸を共有する情報源にも関わらず、その解像度、発生時間間隔などが大きく異なるという、情報の「不均質さ」から生じる次の問題の解決を目的とした。

- (1) データの時空間的なばらつきが大きいいため、通常データベース検索手法では意図に沿った結果が得られない。
- (2) センサデータの空間的な解像度や発生時間間隔などに大きな偏りがあるため、組み合わせるとばらつきが大きすぎて一体的な表示や対話的な操作・統合が難しい

(3) 不均質な情報源はそのメタデータも不均質になりやすく、適切に記述・整備することが困難である。

3. 研究の方法

この問題解決に対し、以下の4つの要素課題を考え、それらを最終的に統合し、標準化とあわせて成果を普及することとした。



(図2 研究の方法)

各要素課題は、方式の研究開発にとどまらず、実際のデータを使ったプロトタイプ実装作成や実験を行うこととし、実用性に基づいた評価を行うこととした。得られた手法や技術については、関連コンソーシアムでの活動を通じて標準化の活動を行うことで、普及を図ることとした。

(1) 不均質さに着目した「データ統合基盤」

不均質な時間空間センサデータに対する検索では、解像度や時間幅などの検索条件にばらつきが大きいため、利用者の念頭においた時間的・空間的な指標と大きく異なる結果が得られる可能性が高い。例えば、ある1時間に発生したデータを検索したとき、処理不可能ほどの膨大な数の結果が得られたり、逆に1件も結果が得られなかったり、という状況が多々発生する。また、1時間程度の誤差でも許されるため厳密に『=』で扱う必要のない場合など、結果の許容範囲にも違いがある。空間的にも、ある範囲を含む画像という検索に対し、解像度の荒い地球全体の画像を戻すことは意味がなく、意図に合わせた結果を返す必要がある。

この問題を解決するために、情報検索における索引手法や検索条件の緩和等に基づいた大規模検索基盤の研究開発を行なった。

(2) Web ブラウザに基づく「対話的検索と可視化」

本研究の応用では、人間の処理が介在することが極めて多く、例えば『衛星画像中の興味地点が雲に覆われているか』という情報は、画像中の雲量の%だけで分かるものではなく、利用者が目で見たと判断を下さざるを得ない。Web ブラウザのような対話的な環境

での表示・操作が不可避である。しかし、ペタバイト級の不均質なセンサデータを対象とする場合、メタデータだけでも膨大な量となり、有効な可視化手法、検索手法が存在しない。また、異種センサデータを統合する事を考えると、対話的処理同士の統合・連携をも考える必要がある。そこで不均質なデータの可視化とそれに基づく効果的なナビゲーション検索を実現するため、従来から行ってきたブラウザにおける高速 XML 転送やマッシュアップの研究に基づき、ユーザとの対話性を重視した大規模なメタデータへの効率的な問い合わせとナビゲーション についての研究を行った。

(3) 不均質センサデータに対する「ストリーム処理基盤」

研究分担者らは、従来型のデータベースや Web 等のみではなく、センサ、位置情報源等のストリーム情報源をも統合対象とする情報統合技術を研究開発してきた。本統合システムは、リレーショナルモデルをベースとした基本統合機能を有するが、ビデオや音声等の連続メディアの処理にも適用可能である。また、応用 目的向けのプログラムを外部関数として追加可能としている。これによって、リアルタイムセンサデータのような時系列データの類似検索や、サーベイランスカメラ映像ストリームに対する解析機能を組み合わせた情報統合等を可能としている。これを発展させ、ペタバイト級の大規模衛星データ等の静的センサデータと動的で実時間性の高いセンサデータの統合利用を実現すべく研究を行った。

(4) XML に基づく「メタデータの記述と利用」

これまでにも、数多くの研究プロジェクトにおいて、センサデータのためのメタデータ記述、あるいはメタデータに基づくセンサデータ検索に関する検討が行われてきており、これらの多くは近年の XML の普及に伴い、標準的なフォーマットして採用している。この状況で、ペタバイト級の不均質なセンサデータの統合利用のために、分散 XML メタデータベースの構築と高速な検索やメタデータを利用した、時空間センサデータ問合せの効率化を研究した。

(5) 統合利用基盤と「標準化」への検討

本研究テーマで開発される個々の要素技術を統合、連携させて実用的な基盤を実現するための研究開発を行った。あわせて、実際の環境における実験や応用による評価を行って有用性を検証した。また、最終的に得られた技術仕様を、OGC, OGF など標準化関連の議論に反映するための活動を行った。

4. 研究成果

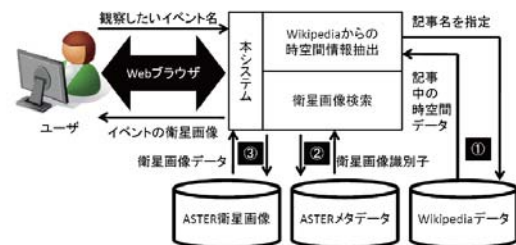
要素研究の成果は最終的に大きく 2 つに分けられ、統合化・応用・実用化と標準化とあ

わせ以下のようにまとめることができた。

(1) 衛星画像と Web 情報の統合利活用技術：3 節の課題及び方法 (2) と (3) を結びつけ、衛星センサのストリーミング情報と、ブラウザを使った対話的検索等で得られる Web 情報の統合的な利用技術を研究開発した。具体的には、地域開発や建築物の建築等センサ情報の差分から得られる情報と、Web で得られる事件等の情報を結び付け世の中の事象(イベント)の正確な理解を支援できる技術を開発した。

① Web 情報からのイベント抽出

本技術項目では集合知から観測知(衛星画像)の検索手法を研究した。衛星画像を検索するためには主に時刻情報(年月日)および空間情報(緯度および経度)が必要であるが、これらを集合知から得るシステムを開発した(発表④)。



(図3 Web 情報からのイベント抽出システムの構成)

具体的には、ユーザが衛星画像を閲覧したい事象についてのキーワードを入力し、Wikipedia 記事を検索する。次に、その Wikipedia 記事中の時空間情報を抽出し、ユーザに提示する。最後に、抽出された時刻情報からユーザが衛星画像を閲覧したい記事中に存在するイベントを選択することにより、衛星画像によるイベントの時空間的影響を閲覧することができるようになった。

これは、Wikipedia の記事に主要な出来事が年表のように記載されているセクションが存在する(年表セクションと呼ぶ)点に着目して時系列の抽出を行った点が特徴であり、時刻情報が時系列順に最も多く出現するセクションを年表セクションとすれば時系列を抽出できること示し、アプリケーションフレームワーク(雑誌②)上で示した。

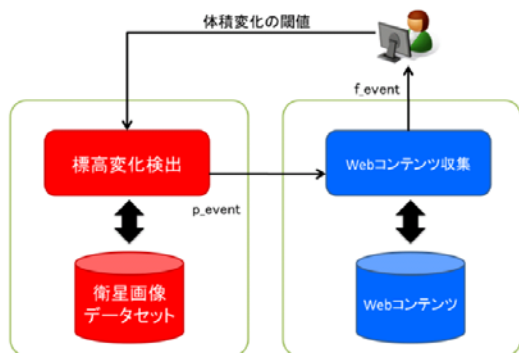
図4の例は、「静岡空港」の記事から年表セクションを求めた結果であり、「本体工事に着手した日」、「開港日」など、静岡空港の建設の様子を知るための必要な時刻情報が得られ、これに基づく時系列的な衛星画像には建設の様子が示されており、時空間的な広がりのあるイベントをより正確に理解できるようになった。



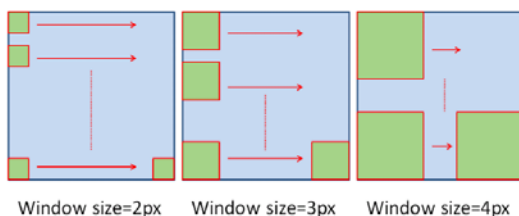
(図4 Wikipediaの情報と対応する衛星画像)

② 衛星画像からのイベント抽出

本技術項目では、逆に衛星画像の時系列に基づく差分検出をイベントとして、web情報と関連付ける方法を研究開発した。具体的には、衛星センサからの3次元標高モデルに対して探索ウィンドウを用いた領域検出手法を提案し、従来に比べて100倍高速な変化抽出を可能とした(発表⑤)。



(図5 衛星画像からのイベント抽出システム)



(図6 探索ウィンドウによる領域検出手法)

さらに、時系列的に離散的な衛星画像の差分から、標高変化の発生開始と終了の時点の予測を行う方法を開発した。

これにより検出された標高変化の発生領域や発生時間に対して逆ジオコーディング

を行い、建造物など得られた名称をキーとしてWeb検索を行い、コンテンツ集合を収集する。逆ジオコーディングに用いた市町村名をあわせて検索することで、同一名の異なる建造物との混同を避けられる。

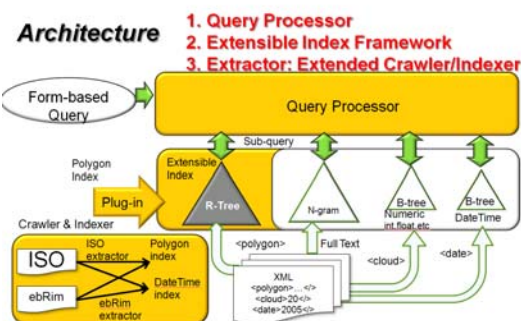
さらに、入力した標高変化イベントの発生時刻により近い情報が記述されているWebコンテンツが高スコアとなるように、収集したコンテンツ集合に対してリランキング処理を提案して精度を向上した。これは、収集したコンテンツ集合の各コンテンツを解析し、コンテンツに記述されている時間表現を抽出し、より妥当なイベントを上位にランクするものである。

以上のような①、②の2つの相補的な技術研究開発により、Web情報と衛星画像の統合利活用が可能になった(発表③)と考える。

(2) メタデータ統合検索基盤:

これは、3節の課題及び方法の①、④を結び付けるもので、大規模な異種のメタデータ、それもXMLベースのデータに対して効率的な検索基盤を提供するものである。基本的な問題は、特にGUIなどで意図しない個数の結果が返るため、結果の個数に回答時間が影響されにくい検索基盤を提供することと、XMLベースの様々な異種データフォーマットを統一的な方式で管理できることである。

このために本研究では、全文検索システムの技術を利用して、利用者が拡張できるデータ型索引にもとづく検索システムを開発し、この上にOGC(Open Geospatial Consortium)のカatalog検索規格を実装した(発表⑥)。これは関係データベースに基づく従来の実装に比べて特に検索結果が多い場合に大幅に有利と案。一方、データベースの異種性については、全文検索とタグの検索、データ型に応じた条件検索を組み合わせることで広範囲なデータフォーマットに対応できるようになり、後述のような応用に利用することができた。



(図7 拡張できる索引による情報検索システムの構成)

メタデータの表現については、RDF(Resour

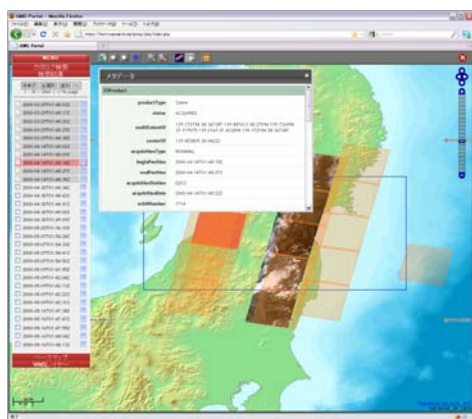
ce Description Framework) と呼ばれるセマンティック Web の分野で標準的なメタデータ形式を扱い、その効率的なアクセス手法(発表⑦)や、メタデータが分散している環境の上での効果的なメタデータ発見、モニタリング、更新波及を支援する手法などを研究開発し、メタデータの分散化に対応できる技術を確立した(雑誌③、⑤)。

(3) 応用と実用化、標準化

① 応用: 応用については、気象データの分析とアクセス方法の研究に行い、サポートベクターマシンと隠れマルコフモデルによる機会学習に基づいた自動判別技法を提案して気象パターンを判別する検索システムのプロトタイプを開発し、実データを評価すると共にアンケート調査等で検証して有効性を確認した(雑誌④)。

さらに、GUIを持つアプリケーションなど、ブラウザを基礎とした応用についてのサービス連携フレームワークを研究開発し、HTML 5 に基づいて既存の応用を改造することなく連携する技術を開発し、これまで示した個々の要素技術の統合を可能にする(発表①、②)と共に、超高解像度の画像の表示のためのブラウザ技術を開発した(雑誌①)。

② 実用化: メタデータ検索システムについては、GEO Grid における ASTER メタデータカタログに適用してサービス提供を行っている。(発表⑧)



(図8 メタデータ検索システム)

これは、約150万件程度の衛星データに対するメタデータカタログであり、OGCのカタログサービスの規格に準拠している。GUIを用いた検索のため、答えの数が意図しない膨大な数になることが多いが、商用のものも含め関係データベースに基づく実装では応答時間の劣化が大きかった。本システムの導入により、十分な応答時間を達成し、実用的なサービスとして利用されている。また、全文検索を基礎とすることで、地球観測に関わる様々な標準フォーマット(ISO19115やebRIM

Earth Observation profile など)を统一的に扱うことが可能となった。また、Web情報と衛星情報の統合に関しては、衛星データの利用促進に有効な技術であり、今後の実用化を図っていく予定である。

③ 標準化: 標準化については、特にメタデータ検索基盤に対してOGCのカタログサービス(Catalog Service)のワーキンググループにおいて、全文検索、特にOpenSearchと呼ばれる規格に基づく検索仕様の標準化の推進と実装の支援を行った。

以上のように、本研究課題について当初設定した目的や課題がおおむね解決でき、これによって不均質な情報源の相互連携と管理が促進されて知見の創出に有益な成果が得られたと考える。また利用できるサービスを提供できたことで、実用性についても実証できたと考えられる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計9件)

- ① 横山昌平, 石川博, Webブラウザによる超高解像度可視化基盤の開発, 情報処理学会論文誌, 査読有, Vol. 52, No. 1, 2011, pp56-67
- ② 横山昌平, 黒井星良, 岡本章裕, 石川博, WebGIS アプリケーションフレームワークrinzo.maの設計と応用, 情報処理学会誌データベース, 査読有, Vol. 3, No. 1, 2010, pp82-95.
- ③ M. E. Gutierrez, I. Kojima (CA), S. Mirza, O. Corcho, A. Perez, Accessing RDF(S) Data Resources in Service-based Grid Infrastructure, Concurrency and Computation, Practice and Experience, 査読有, Vol. 21, No. 8, 2009, pp1029-1051
- ④ 木村広希, 川島英之, 北川博之, サポートベクターマシンによる揮発配置の自動分類, 日本データベース学会論文誌, 査読有, 2009, Vol. 8, No. 1, pp59-64
- ⑤ S. Mirza, I. Kojima (CA), S-MDS: Semantic Monitoring and Discovery System for the Grid. Journal of Grid Computing, 査読有, 2009, Vol. 7, No. 2. pp205-224

〔学会発表〕(計37件)

- ① S. Yokoyama, H. Ishikawa, Parallel Distributed Rendering of HTML5 Canvas Elements, The 11th International Conference on Web Engineering (ICWE), 2011.06, Cyprus (to appear).
- ② A. Matono, A. Nakamura, I. Kojima, A. Mas

- hup Tool for Cross-Domain Web Applications using HTML5 Technologies, The 13th, Asia Pacific Web Conference (ApWeb), 2011. 04. 05, Beijing, China.
- ③ 岡本章裕, 横山昌平, 川島英之, 天笠俊之, 的野晃整, 石川博, 北川博之, 小島功, 衛星画像とWeb情報の統合的な利活用手法の提案, データ工学研究会, 2010. 12. 06, 東京
 - ④ A. Okamoto, S. Yokoyama, N. Fukuda, H. Ishikawa, Proposal of Spatiotemporal Data Extraction and Visualization System Based on Wikipedia for Application to Earth Science, 9th International Conference on Computer and Information Science, 2010. 8. 19, Yamagata, Japan.
 - ⑤ T. Kakagi, H. Kawashima, T. Amagasa, H. Kitagawa, Providing Constructed Buildings Information by ASTER Satellite DEM Images and Web Contents, Data Intensive eScience Workshop, 2010. 04. 04, Tsukuba, Japan.
 - ⑥ I. Kojima, M. Kimoto, A. Matono, OGC Catalog Service for Heterogeneous Earth Observation Metadata using Extensible Search Indices, 6th ACM Workshop on Geospatial Information Retrieval, 2010. 02. 18, Zurich Switzerland.
 - ⑦ I. Kojima, M. Kimoto, Implementation of a Service-Based Grid middleware for Accessing RDF Databases, SEMELS Workshop, On the Move Federated Conference, 2009. 11. 02, Vilamoura, Portugal.
 - ⑧ I. Kojima, Data Management in GEO Grid(Invited), EGEE 09 Conference, 2009. 09. 24, Barcelona, Spain.
 - ⑨ S. Yokoyama, I. Kojima, H. Ishikawa, FREDDY: A Web Browser-friendly Lightweight Data-Interchange Method Suitable for Composing Continuous Data Streams, Composable Web' 09 Conference, 2009. 06. 23, San Sebastian, Spain.
 - ⑩ S. Lynden, S. Mirza, I. Kojima, Service-based Data Integration using OGSA-DQP and OGSA-WebDB, International Conference on Grid Computing, 2008. 10. 1. Tsukuba, Japan.

[その他]

ホームページ等：なし

6. 研究組織

(1) 研究代表者

小島 功 (KOJIMA ISAO)

独立行政法人産業技術総合研究所・情報技

術研究部門・研究グループ長
研究者番号：00356982

(2) 研究分担者

的野 晃整 (MATONO AKIYOSHI)

独立行政法人産業技術総合研究所・情報技術研究部門・研究員

研究者番号：10443227

北川 博之 (KITAGAWA HIROYUKI)

筑波大学・大学院システム情報工学研究科・教授

研究者番号：00204876

天笠 俊之 (AMAGASA TOSHIYUKI)

筑波大学・大学院システム情報工学研究科・准教授

研究者番号：70314531

川島 英之 (KAWASHIMA HIDEYUKI)

筑波大学・大学院システム情報工学研究科・講師

研究者番号：90407148

石川 博 (ISHIKAWA HIROSHI)

静岡大学・情報学部・教授

研究者番号：60326014

横山 昌平 (YOKOYAMA SHOHEI)

静岡大学・情報学部・助教

研究者番号：20443236