粒度の異なる多様な時空間データの保存・抽出システムの設計と実装

鷲田 健斗[1]　廣井 慧[3]　米澤 拓郎[1]　河口 信夫[1,2]

概要：近年では，スマートフォンに代表されるモバイル端末の普及やセンサ技術の進歩により，多くの種類の時空間データが得られるようになった．これに伴い，これらのデータの活用のために時間や空間の粒度の変換を行うことの需要が高まってきた．しかし，データの時間粒度，空間粒度が異なる複数のセンサに関してはデータの統一的な扱い方が十分に検討されていない．このままでは，災害時の情報共有や地域のマーケティングなどのデータ交換プラットフォームで円滑にデータを共有できない．本論文では，粒度の異なる時空間データの保存・抽出が可能なシステムとして，統一的なデータ形式で扱うことを考え，そのデータ形式への変換，データ形式に対するデータの取り出し方を提案する．

Design and Implementation of Various Spatio-temporal Data Storage and Extraction Systems with Different Granularities

KENTO WASHIDA[1]　KEI HIROI[3]　TAKURO YONEZAWA[1]　NOBUO KAWAGUCHI[1,2]

1. はじめに

近年ではスマートフォン等のモバイル端末の普及やセンサ技術の進歩により，多くの種類の時空間データが得られるようになった．これらのデータの活用のために時間や空間の粒度変換への需要が高まり，ここ 10 年ほどで時間粒度のみの変換や空間粒度のみの変換は容易に行えるようになってきた．また，時間と空間の両方の粒度を考慮した可視化として Kompreno(人流)[1]，混雑度マップ [2]，XRAIN[3] などがある．粒度の変換でデータがより扱いやすいものになり，分析をより簡単に行うことができる．

しかし，このような可視化の事例は単一ないし類似する時空間データのみを対象としている．これはセンサによってデータの時間と空間の両方の粒度が異なるデータの統一的な扱い方が十分に検討されていないためである．センサ毎に分析の仕組みを作ることは効率的ではなく，複数の仕組みが存在する状況ではユーザの負担も増加する．複数種類のセンサが利用される状況として，災害時の情報共有や地域のマーケティングなどのデータ交換プラットフォームでのデータ共有などが考えられる．このような状況で円滑なデータ共有を可能にするために複数種類のセンサのデータを統一的に扱えるようにする必要がある．

時空間データを扱う研究として，特定のセンサに対するデータの扱い方，分析や特定の状況での利用をしている研究が多くある．また，特定の状況下ではなく広い範囲を対象としている研究は概念レベルのものがほとんどであり，今回のような想定である多様な時空間データに対応できる正確なモデルは存在しない．本論文ではユーザが同じ操作で多様な時空間データを扱えるようにするため統一的なデータ形式を考え，そのデータ形式への変換，データ形式に対するデータの取り出し方を提案する．本研究室では時空間データの中でも人流データが得られる WiFi センサ，GPS センサ，通過センサの 3 つのセンサを扱っているので，それらのデータを対象として時空間粒度変換手法を検討した．

2. 関連研究

多粒度の時空間データの粒度変換に関して，時空間オブ

[1] 名古屋大学大学院工学研究科 Graduate School of Engineering, Nagoya University
[2] 名古屋大学未来社会創造機構 Institutes of Innovation for Future Society, Nagoya University
[3] 京都大学防災研究所　Disaster Prevention Research Institute, Kyoto University

ジェクトモデルを考えた粒度変換の提案が存在する [4], [5]. このモデルは粒度の関係について深く考察されているがまだ概念レベルであり，多くの課題を残していて実用化はされていない．また，統計的なデータを対象としているのでセンサデータを対象とする我々の目標とは異なる．多粒度の時空間データの統合に関する研究として共通の最適な粒度への変換のフレームワークが提案されているが，自由な粒度への変換はサポートされていない [6]．この研究も時空間の粒度変換を扱っているが方向性が異なる．

時空間の情報を持つ異種データベースを統合する研究も行われていて，時空間的な関連性を評価してその関連性を元にデータを統合している [7], [8]．これはある程度似ているデータに対しての研究であり，多様な時空間粒度のセンサデータを対象に考えると複数の形式が存在するため関連性を評価するのは難しく，統合されたデータの形式も一定では無い．また，時空間データの可視化の検索支援機構が提案されている [9]．この研究では単一のモバイルセンサデータが対象とされており，そのまま使うことはできないがよりわかりやすい可視化として参考にしたい．以上のように時空間に関する研究は広く行われているが複数の時空間粒度の異なるセンサデータを対象としている研究は多くない．

3. 設計方針

本章では，本研究で提案する多様な時空間データの保存・抽出システムの設計について述べる．時間・空間的な変化が生じ，かつ同様の推定ができるにもかかわらず利用するセンサで得られるデータ形式が異なるものとして，人流データを例に挙げて設計方針を考える．実際の人流データには次のようなものがある

- GPS センサ (スマホ):時間，スマホ ID，緯度経度，その他メタデータ
- WiFi パケットセンサ:時間，センサ ID，電波強度，その他メタデータ
- 通過センサ:時間，センサ ID，エリアへの出入，その他メタデータ

データを自由な粒度で取り出すためには出力されるデータが一定の形式である必要がある．データを一定の形式にして扱うためには以下の 3 点を考える必要がある．

- データベースに保存されるデータの形式
- センサから得られたデータの保存される形への変換
- データベースから任意の粒度で取り出す仕組み

これらをそれぞれ標準データ形式，データの保存，データの抽出として分けて考える．

3.1 標準データ形式

標準データ形式は時間と空間の粒度が自由に変換できるものでなければならない．時間粒度が簡単に変換できるものとして時系列データがあげられる．続いて，空間粒度が簡単に変換できるものとしてメッシュがあげられる．以上のことから，メッシュそれぞれが時系列データを持つようにすればメッシュ単位での時空間粒度の変換が可能になる．このデータ形式のイメージが図 1 である．地図中の各メッシュには，それぞれ対応する時系列データが存在し，例えばメッシュ 1 と 2 では図下側の折れ線グラフのようなものとなる．このデータ形式は空間粒度が単純な形のため扱いやすく，メッシュの組み合わせにより任意の形も作成することができる．一方でメッシュをまたがるデータを表現するのには工夫が必要であり，そこには限界が存在するという短所もある．

図 1　標準データ形式のイメージ図

表 1　センサの種類

種類	時間粒度	空間粒度
GPS センサ	即時	ポイント
WiFi センサ	即時	一定の範囲
通過センサ (出入り口全て)	即時	仕切られた範囲

3.2 データの保存

センサから得られたデータを 3.1 節で決めたデータ形式にあわせるためにメッシュ毎のデータに変換する．想定されるデータは表 1 に示すように時間粒度は即時である．空間粒度について GPS センサはポイント，WiFi センサと通過センサはある広さの範囲でありこれはポリゴンである．これらの粒度を想定することで幅広いセンサに対応できると考えられる．ここでいう即時とはアクション (アプリを起動する，特定の場所を通る) を起こしたとき，そのタイ

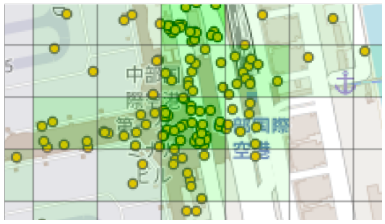

図 2　ポイント

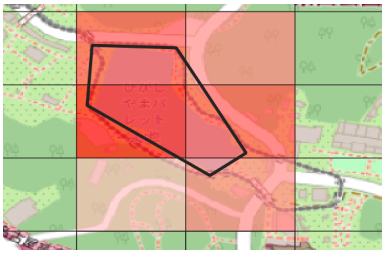

図 3　ポリゴン

ミングでセンサがデータを得る場合のことである．ポイントは緯度経度，ポリゴンは建物内などの情報のことである．時間粒度に関して基本的にセンサは記録された時刻のデータを持っているのでその時刻を元に時系列データを作成する．空間粒度に関しては，ポイントデータとポリゴンデータで異なる扱いが必要になる．ポイントのデータは図 2 のようにメッシュ内のデータを 1 としてカウントし，ポリゴンは図 3 のように面積按分をしてそれぞれのメッシュにふりわける．これらの操作によりセンサから得られたデータを 3.1 節で決めた標準データ形式に変換する．

3.3　データの抽出

ユーザが任意の粒度でデータを得られる取り出し方を考える．時間粒度は開始時間，終了時間，時間幅で指定する．空間粒度は図 4 のように指定した 2 点を対角の頂点に持つ長方形で示される範囲と，図 5 のようにポリゴンで指定された好きな範囲の 2 種類を扱えるようにする．ポリゴンに関して返すデータはポリゴンと重なるメッシュのデータの合計とする．これらを基本的な抽出機能とする．

また，時間と空間の 2 つの粒度を考える必要があるのでユーザがデータを正確に把握していない場合，適当な粒度を探すのが困難であることやピーク位置の推移を知りたい場合が想定される．このようにユーザの要望を想定して，応用的な抽出機能を追加してより使いやすく便利なものにする．

4.　実装

3 章で述べた設計に基づき，図 6 に示す構成で実際のシステムを Python3.6.0，PostgreSQL12.2，PostGIS3.0 を使用して実装した．システムはセンサからデータを受け取り統一のデータ形式にする保存機能，変換されたデータを蓄えておくデータベース，ユーザの要求により必要なデータ

図 4　2 点で指定する範囲

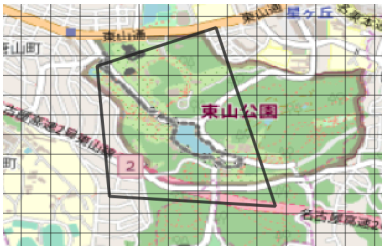

図 5　ポリゴンで指定する範囲

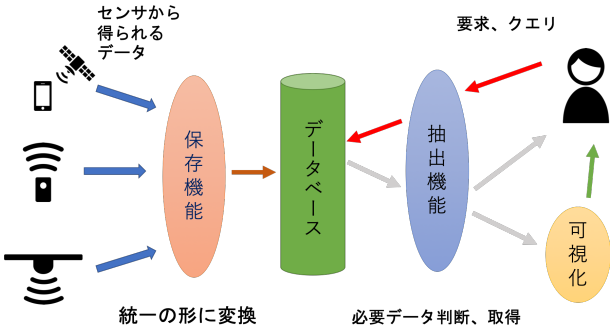

図 6　システム構成

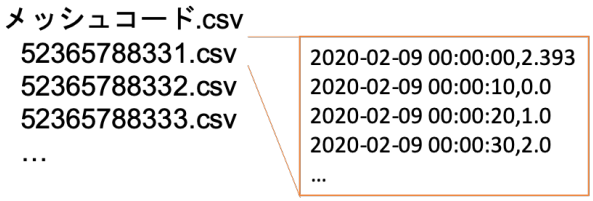

図 7　csv によるフラットファイルデータベース

を判断しデータベースから持ってくる抽出機能によって構成される．このとき粒度変換のために必要な計算が増えるが，できるかぎり動作が遅くならないように実装する．

4.1　データベース

このシステムではデータベースに大量のデータが置かれるのでデータベース内で計算を行うことは避ける．するとデータベースはデータを置くだけになり，非常にシンプルな実装にできる．扱いやすさから，メッシュコードをファイル名に持ち，中身が時系列データの csv ファイルである図 7 のようなフラットファイルデータベースを使う．これならば，メッシュコードで管理が可能になり，余計なデー

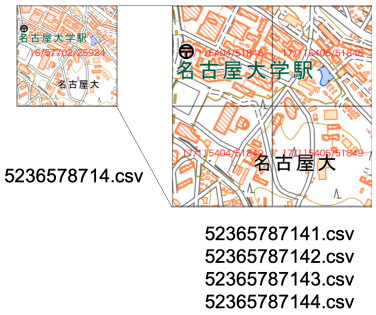

図 8　異なるサイズのメッシュへのデータ分配

図 9　保存の様子

タを読まずに抽出を行える．また，抽出の際に，毎回小さいメッシュの情報を用いて大きいメッシュの情報を計算することは計算量的にも時間的にも負荷が大きいため，データ保存時に図 8 のように各粒度のメッシュに分配する．

4.2　保存機能

保存する際は一定の時間幅ごとに得られた値の集計を行い，頻度を一定に保つ．これはセンサのデータを得る頻度が高くデータが膨大な量になることを防ぐためである．今回は人数を対象としており，かつ統計的に計算するので小数点も考慮し以下の処理を行う．ポイントのデータに関しては，緯度経度からメッシュコードを調べ，該当メッシュの該当時刻のデータに+1 する．ポリゴンのデータに関しては，面積按分をおこない該当メッシュの該当時刻のデータに面積比を加算する (図 9)．面積按分は PostGIS にあらかじめメッシュコードを持つメッシュのジオデータを入れておき Python で SQL 文を生成して行う．この SQL 文でメッシュコードと面積比を取得する．データの保存を 1 行毎に行うと処理時間が長くなるので，大量のデータを投入するときは先に全てのデータを処理して一時的な加算のデータを作り，まとめてファイルと統合する．リアルタイムのデータに対してはまだ実装できていないが，まとめられた時間粒度のタイミングで統合を行う．

4.3　抽出機能

指定された範囲がどのメッシュと重なっているかは保存機能と同様に PostGIS と SQL を用いておこなう．基本的な抽出機能に関して，2 点の緯度経度の指定は，図 10 のように長方形のポリゴンを作成し重なるメッシュを調べてそれらを指定された時間幅でまとめて返す．ポリゴンの指

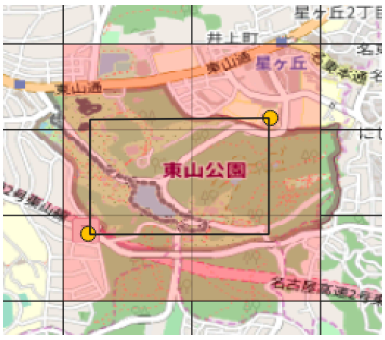

図 10　2 点と重なるメッシュの判断

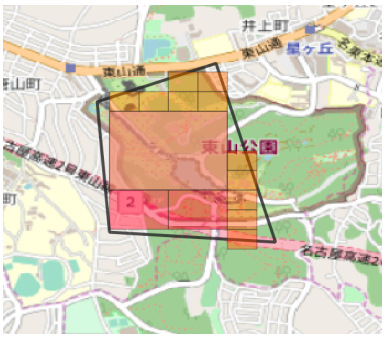

図 11　ポリゴンと重なるメッシュの判断

表 2　保存・抽出の関数仕様

データの保存 (上:ポイント　下:ポリゴン)

store_data_by_point(time, lat, lon, timerange)
store_data_by_polygon(time, polygon, timerange)

データの抽出 (上:2 点の指定　下:ポリゴンの指定)

retrieve_data_by_2points(lat1, lon1, lat2, lon2, starttime, endtime, timerange, meshsize)
retrieve_data_by_polygon(polygon, starttime, endtime, timerange)

定は，図 11 のように指定されたポリゴンと半分以上重なるメッシュを合計した値を指定された時間幅でまとめて返す．このとき，最小サイズのメッシュで作るのではなく，大きいメッシュから重なっているかを調べ計算の負担を減らす．応用的な抽出機能として，ピーク位置の推移がわかる機能を提案する．この機能は，各時間毎に最大値のデータが存在するメッシュを探しそのメッシュコードを返す．また，自動粒度変換機能を提案する．この機能は，ユーザに指定された対象エリアとそのエリア属性を元に，ユーザが必要としている空間粒度を推定し，適切なメッシュサイズを提案する．今回は自動粒度変換機能は実装できていない．

4.4　関数仕様

4.1〜4.3 節の機能を表 2 に示す関数として実装した．具体的にはデータ保存 (store_data) 時は時間，時間幅，ジオメトリの情報を使って統一のデータ形式へ変換する。ジオメ

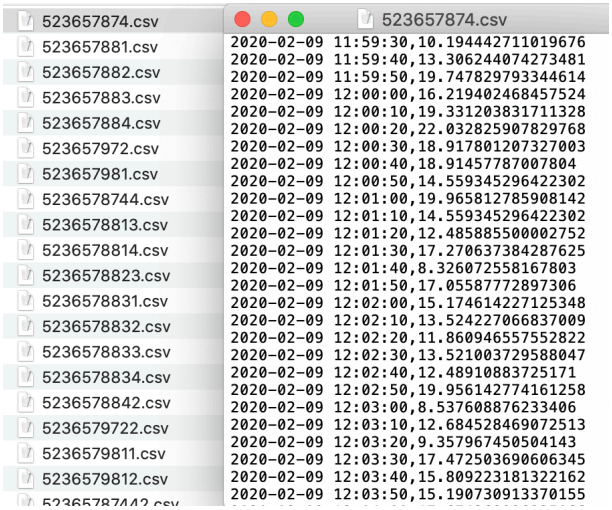

図 12　WiFi パケットセンサの変換されたデータ

トリの情報は、ポイントとポリゴンに対応していてどちらでも登録することができる。また、データ抽出 (retrieve_data) 時はジオメトリ，データを取得する開始・終了時間，時間幅，メッシュサイズ (2 点のみ) の情報を使ってデータを取り出す．抽出に関してはメッシュサイズが指定されない場合，大きいサイズから範囲内のメッシュの数を調べ，一定数になる粒度でデータを返す．

5.　評価

本章では実際に WiFi パケットセンサ，GPS センサを使って，システムを動作させる．対象とする WiFi パケットセンサは東山動植物園に設置されており，休日の 12 時台に数十万レコードほどのデータが得られる．データは「タイムスタンプ，AMPID，HMACID，OID，Flag，rssi，Counts」のフィールドを持っている．このデータに対し，保存のために空間的な有効範囲，まとめる時間幅を指定する．有効範囲については半径 50m の円，まとめる時間幅は 10 秒とした．指定したものとタイムスタンプから図 12 のようなデータに変換される．抽出の時に，2 点 ((35.153929, 136.976423)，(35.159192，136.985178))，6 次メッシュ，取得時間 2020-02-11 09:00:00 ～ 17:59:59，1 時間幅で指定するとそれに応じた結果が得られる (図 13)．この結果の 14 時，15 時，16 時を QGIS で可視化したものが図 14 である．

対象とする GPS センサは特定のアプリユーザの位置情報を取得するものであり，セントレアエリアを対象とした範囲でデータが得られ，12 時台に 3，4000 レコードほど得られる．データは「タイムスタンプ，hashid，緯度，経度，accuracy，os」のフィールドを持っている．このデータに対し，保存のためにまとめる時間幅を指定する．今回の場合はまとめる時間幅は 10 秒のようにした．指定した時間幅とタイムスタンプ、緯度経度により図 15 のようなデータに変換される．抽出の時に，POLY-

	52365798111	52365788333	52365788331	52365788313
2020-02-11 09:00:00	6.262369	15.510751	9.381198	0.764407
2020-02-11 10:00:00	13.660029	30.866473	18.731502	2.972406
2020-02-11 11:00:00	18.431559	39.933102	22.708542	5.001293
2020-02-11 12:00:00	27.764326	55.050021	30.784834	5.362407
2020-02-11 13:00:00	32.811435	57.908017	27.627751	4.683223
2020-02-11 14:00:00	30.485416	56.929352	27.300365	4.442865
2020-02-11 15:00:00	26.678089	46.932006	19.426121	3.394477
2020-02-11 16:00:00	11.877540	25.099326	8.686883	1.593526
2020-02-11 17:00:00	1.462212	1.656456	0.049241	0.017045

図 13　2 点の指定による抽出の様子

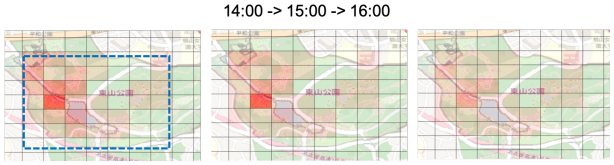

図 14　QGIS による可視化

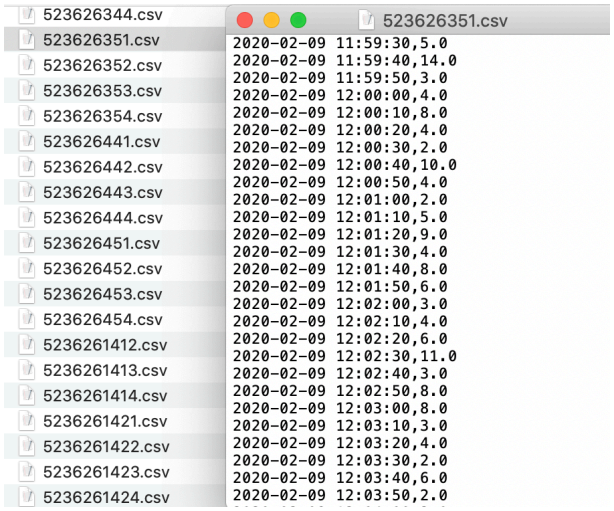

図 15　GPS センサの変換されたデータ

GON((136.81483 34.86621, 136.80554 34.86518, 136.80572 34.85124，136.81735 34.85158，136.81483 34.86621))，取得時間 2020-02-11 11:00:00 ～ 15:59:59，30 分幅で指定するとそれに応じた結果が得られる (図 16)．

このようにして空間粒度がそれぞれ円 (ポリゴン)，ポイントのデータに対して統一的なデータ形式に変換しさまざまな時空間粒度で抽出することに成功した．

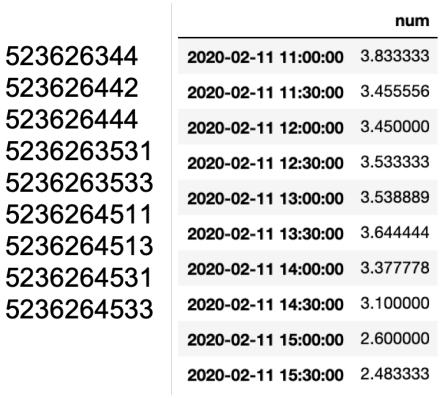

図 16 抽出機能で必要と判断されたメッシュ (左) と結果 (右)

6. まとめ

本論文では，粒度の異なる時空間データの保存・抽出システムを設計，実装し実際のデータを用いて動作させた．設計方針では，人流に関するデータが得られる複数種類のセンサを想定して，標準データ形式・保存・抽出を定義した．実装では，標準データ形式をファイル名にメッシュコードを持つ csv ファイルで表し，異なる空間粒度に対応できる保存機能と好きな粒度で取り出せる抽出機能の動作を提案した．このときデータベースで複数のサイズのメッシュデータを用意することで抽出時の計算にかかる時間を減らしている．そして実際の WiFi パケットセンサのデータと GPS センサのデータを対象にシステムを動作させ，様々な粒度への変換が可能であることを確認した．

本論文の課題として，ユーザインターフェースと可視化といった実際にシステムを利用するための部分が未実装である点，蓄積されたデータのみにしか対応していない点，応用的な抽出機能が少ない点があげられる．ユーザインターフェースと可視化の実装について，これらの実装は実際に使うために必要不可欠であり，有用性を評価するためにも必要である．蓄積されたデータのみにしか対応していない点について，通信技術の進歩により多くのリアルタイムのデータが得られることが想定されるのでリアルタイムのデータも扱えるようにする必要がある．応用的な抽出機能について，現時点ではデータが大きく増減したエリアがわかる機能を実装しようと考えている．これにより何かの出来事が起こったエリアがわかるようになる．

今後，本研究室で開発されている Synerex[10], [11] との連携を考えている．Synerex とは需給交換プラットフォームであり，これを参照することで複数のデータソースへのアクセスやリアルタイムのデータの取得が可能になる．また同じく本研究してで開発されている時空間データ可視化ライブラリである Harmoware-VIS[12] を用いた可視化も実現できる．

謝辞

本研究は，JST CREST JPMJCR1882，NICT 委託研究，科研費基盤研究 B(17KT0082) の支援を受けたものです．

参考文献

[1] Kompreno https://www.agoop.co.jp/kompreno/
[2] 混雑度マップ https://lab.its-mo.com/densitymap/
[3] XRAIN 全国概況画面 http://www.river.go.jp/x/xmn0107010.php
[4] Bertino E.，Camossi E.，Bertolotto M. :Multi-granular Spatio-temporal Object Models: Concepts and Research Directions. In: Norrie M.C.，Grossniklaus M. (eds) Object Databases. ICOODB 2009. Lecture Notes in Computer Science，vol 5936. Springer，Berlin，Heidelberg (2010)
[5] Bertino E.，Cuadra D.，Martínez P. :An Object-Relational Approach to the Representation of Multi-granular Spatio-Temporal Data. In: Pastor O.，Falcão e Cunha J. (eds) Advanced Information Systems Engineering. CAiSE 2005.Lecture Notes in Computer Science，vol 3520. Springer，Berlin，Heidelberg (2005)
[6] M. Chen，S. Gao and X. S. Wang :Converting spatiotemporal data Among heterogeneous granularity systems, 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver，BC，pp. 984-992 (2016)
[7] 細川 宜秀，石橋 直樹，八代 夕紀子，清水 康 :マルチデータベース環境における時間的・空間的関連性評価によるデータ結合方式，情報処理学会研究報告，Nov. 1999，Vol.40 No.SIG 8(TOD4) (1999)
[8] 石橋 直樹，細川 宜秀，清木 康 :時空間的文脈に応じた動的関連性計量機構を有する異種データベース間結合方式，情報処理学会研究報告，Mar. 2002，Vol.43 No.SIG 2(TOD13) (2002)
[9] 佐崎 悠，神崎 映光，原 隆浩，西尾章治郎 :時空間的なデータ分布とその変化を考慮したモバイルセンサデータの検索支援機構について，DEIM Forum 2015 C2-1 (2015)
[10] 河口 信夫，米澤 拓郎，廣井 慧 :Synerex: 超スマート社会を支える需給交換プラットフォームの設計コンセプトと機能，情報処理学会研究報告，Vol. 2020-UBI-65，No. 49，pp. 1-6 (2020)
[11] Synerex Project https://github.com/synerex
[12] Harmoware https://github.com/Harmoware