

論文 / 著書情報
Article / Book Information

論題(和文)	SIBM - 避難場所情報に対するRDFデータセットベンチマークツール
Title(English)	
著者(和文)	グエン ホアイ ナム, 荒堀 喜貴, 横田 治夫
Authors(English)	Ngyuen Hoai Nam, Yoshitaka Arahori, Haruo Yokota
出典(和文)	第7回データ工学と情報マネジメントに関するフォーラム論文集, , ,
Citation(English)	, , ,
発行日 / Pub. date	2015, 3

SIBM - 避難場所情報に対する RDF データセット ベンチマークツール

NGUYEN Hoai Nam[†] 荒堀 喜貴^{††} 横田 治夫^{††}

[†] 東京工業大学工学部情報工学科 〒152-8552 東京都目黒区大岡山 2-12-1

^{††} 東京工業大学大学院 情報理工学研究科計算工学専攻 〒152-8552 東京都目黒区大岡山 2-12-1

E-mail: [†]namnh.n.aa@m.titech.ac.jp, ^{††}{arahori,yokota}@cs.titech.ac.jp

あらまし 世界中で毎年、災害による損失や市民生活への影響が問題となっている。災害発生時には、避難情報を元に避難することが必要であるが、災害発生後に、避難に関する情報やデータの量は急速に増加する。このため、適切な災害・避難情報の管理が求められる。一方、近年 RDF 形式で記述されたデータが幅広く使用され、災害情報などを含む様々なデータが RDF 形式で保存されることが増えている。RDF 形式を利用すると、情報の細かな管理やアクセス制限を適切に行うことができるが、既存の災害情報の RDF 表現では、被災者個人間の関係や被災者と医者、ボランティアなどの支援者との関係や、どの支援者がどの被災者にどのような作業を行って良いかといった関係を表現できていない。そこで、本研究では、これらの詳細な関係を RDF 形式で記述した災害・避難情報のデータセット SIBM を提案する。SIBM の目的は、避難場所情報や、被災地における被災者-支援者間のデータ共有に応じる災害情報管理システムの評価を可能にする。具体的には、被災者の怪我の程度などの個人情報を始め、被災者間の関係や災害発生位置による避難場所の配置など現実にある情報を再現できるデータセットを生成することと、家族の人を検索するや避難場所の収容量を調査するなどの現実に近い問い合わせを再現できるクエリセットを目指す。

キーワード 避難場所情報, RDF 記述法, ベンチマークツール

1. はじめに

世界中には台風や地震などの災害による損失や人々の生活への影響が多く見られる。災害発生時、避難することは重要な作業である。避難する際、避難場所情報の管理や、避難する作業の関係者情報が多量に増加することがわかる。避難場所情報や関係者情報などを管理・アクセスする際、適切な管理システムが求められる。

一方、近年、様々な新たなデータ記述法が開発されている。その中に、情報を RDF 形式で保存・共有することが増えてくる。RDF 形式を利用すると、情報の細かな管理やアクセス制限を適切に行うことができ、RDF 形式で記述した災害・避難情報を共有することが考えられている [1]。

また、避難場所情報を共有する際、共有範囲を限定すべき情報を暗号化・復号化することが現実の問題である。例えば、避難作業の関係者の中に、被災者個人情報などアクセス権限を被災者個人間の関係を考慮しながらそのようなアクセス制限を適切に行うシステムが必要である [3]。

そして、災害発生時、情報が急速に増加することがわかる。多量なデータ量を暗号化・復号化することが考えられ、そこで、効率的に暗号化・復号化できる情報管理システムが求められる [4]。

そのようなシステムを評価するために、RDF 形式で記述した避難場所情報ベンチマークが考えられ、本稿では、被災者の怪我の程度などの個人情報を始め、被災者間の関係や災害発生位置による避難場所の配置など現実にある情報を再現できるデー

タセットを生成するベンチマークを構成することと、家族の人を検索するや避難場所の収容量を調査するなどの現実にある問い合わせを再現できるクエリセットを用意することを目的とする。

そこで、本稿では、SIBM (Shelter Information Benchmark) - RDF 形式で記述した避難場所ベンチマークを提案し、データ構造に対する問い合わせクエリや現実にある問い合わせシナリオに対するクエリセットを用意することで、避難場所情報管理システムを評価することを可能にした。

2. 関連研究

近年、災害情報の管理に当たって、災害情報の保存・共有がデータ工学の重要な問題とみなしている。そので、Linked Open Data による災害データ蓄積が巨大な革命であると Jens ら [1] が論じた。

災害対象者または全世界の人々が災害情報を WEB 上で共有することが増えていたとわかった。しかし、情報の増加量や時間とともに、それらの情報がよく構造化されていないことがある。そこで、RDF 形式で記述された Linked Open Data というタームが構成化の問題を解決できると Jens ら [1] が論じた。

この研究では、Ushahidi Haiti プラットフォームを使用することで、災害情報を Linked Open Data による公開・共有するための必要条件や環境設計について考察した。そこで、Linked Open Data が災害情報の構成化問題やセマンティック情報交換可能問題を解決できることを判定した。

RDF/OWL データベンチマークに関して、Guo らが [2] で

LUBM - Lehigh University Benchmark が開発された。LUBM では、大学・学科・学部・授業の情報やその関係者（教授・学生など）に対する OWL データの生成するパターン計画が提案され、LUBM が任意なサイズや様々な特徴を持つデータセットを生成することができた。また、LUBM が生成したデータを用いて、RDF データ構造に対するクエリによる様々な RDF ストレージを評価した。

その他、災害・避難情報に含まれる個人情報など共有範囲を限定すべき情報を共有する際、それらの RDF 形式で記述されたデータへのアクセス制御を考慮しながら暗号化する手法について、児玉らが研究されている [3]。災害・避難情報を暗号化し、暗号化されたデータへのアクセス許可がユーザが持つアクセスレベルによって定められた [3] ではアクセス許可に対する情報の暗号化・復号化手法について研究されている。

また、災害に発生する多量情報を暗号化することが考えられ、ダットら [4] が Map-Reduce による多量の情報を効率的に暗号化・復号化する手法について研究している。

表 1 のように、LUBM が提案したベンチマークが、本稿の対象とするデータモデルとの違いがあり、個人情報の詳細や被災者間の関係などを再現しないデータセットである。また、LUBM では、RDF データ構造や情報の推論などに対する問い合わせクエリを用い、多種のデータストレージを評価を行ったが、情報の集約（例えば、大学情報の場合、学生の平均点などの情報を問い合わせるケース）や関係者間の関係を考慮しながら人や場所を検索など避難する際にある問い合わせを再現しない。そのため、避難場所ベンチマークとして使用することができないと判断する。

表 1: SIBM と LUBM の対象情報

対象情報	SIBM	LUBM
個人情報	氏名、生年月日、年齢、性別、住所、所属、携帯番号、メール	氏名、学歴、メール、携帯番号、授業、専門
対象者間の関係	家族関係、友人関係	指導教員関係
災害による状況	病気・怪我情報、避難先	対象外
対象する場所	避難場所：住所、避難場所名、管理番号、災害種別、収容量、位置情報物品、備蓄情報	大学：大学名、学科、学部、授業

3. 前提知識

3.1 RDF について

RDF (Resource Description Framework) は、WWW 上で資源に関する情報を表わすための言語である。タイトル、著者、ウェブ・ページの更新日、ウェブ・ドキュメントの著作権およびライセンス情報、ある共有資源に対する利用可能スケジュールなどのような、ウェブ資源に関するメタデータの表現を特に目的としている。(注1)

RDF は、人間に表示するだけではなく、アプリケーションが情報を処理する必要がある状況を目的としている。RDF は、この情報を表現するための共通の枠組みを提供するため、意味を損なわずにアプリケーション間で情報交換が行える。共通の枠組みであるため、アプリケーションの設計者は共通の RDF パーサや処理ツールを有効利用できる。異なるアプリケーション間で情報交換できるということは、情報が元々作成された以外のアプリケーションでその情報を利用できることを意味する。

RDF は、ウェブ識別子 (Uniform Resource Identifier (URI) と呼ばれる) を使用して事物を識別し、シンプルなプロパティとプロパティ値で資源を記述するという考えに基づいている。これによって、資源を表わすノードとアークのグラフや、そのプロパティと値として、資源に関するシンプルなステートメントを提供できるようになる。

RDF 形式で記述された情報がトリプル (図 1 の A) あるいは RDF グラフ (トリプルの集合) (図 1 の B) で表現できる。

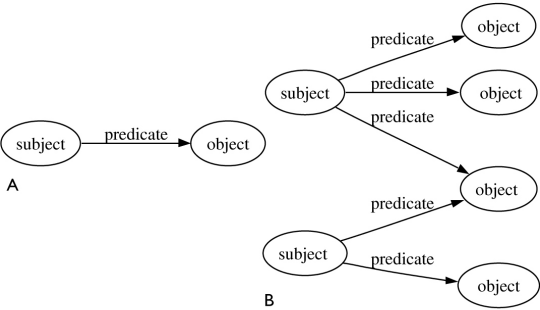


図 1: RDF の例

3.2 SPARQL クエリ言語

RDF に対する SPARQL クエリ言語の構文とセマンティクスを定義している。SPARQL は、データが RDF そのものとして保存されているか、ミドルウェアを通して RDF として見えるのかにかかわらず、さまざまなデータ情報源にまたがるクエリを表わすために使用できる。SPARQL には、必須および任意のグラフ・パターンをその論理積と論理和とともに問い合わせる性能が含まれている。SPARQL は、ソース RDF グラフによる集約、サブクエリ、否定、式による値の作成、拡張可能な値テストやクエリの制約もサポートする。SPARQL クエリの結果は、結果集合または RDF グラフでありえる。(注2)

以下は、SPARQL の例である。図 3 が図 2 の RDF データのクエリであり、そのクエリ結果が表 2 となる。

表 2: クエリ結果

name	mbox
"Johnny Lee Outlaw"	<mailto:jlow@example.com>
"Peter Goodguy"	<mailto:peter@example.org>

(注 1) : <http://www.asahi-net.or.jp/~ax2s-kmtnt/internet/rdf/rdf-primer>

(注 2) : <http://www.asahi-net.or.jp/~ax2s-kmtnt/internet/rdf/REC-sparql111-query-20130321>

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

_:a foaf:name Johnny Lee Outlaw .
_:a foaf:mbox <mailto:jlow@example.com> .
_:b foaf:name Peter Goodguy .
_:b foaf:mbox <mailto:peter@example.org> .
_:c foaf:mbox <mailto:carol@example.org> .
```

図 2: RDF サンプル

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE {
  ?x foaf:name ?name .
  ?x foaf:mbox ?mbox
}
```

図 3: SPARQL サンプル

3.3 Apache Jena と TDB

Apache Jena とは、RDF 形式のデータを処理するアプリケーションを構築するための Java ライブラリである。RDF 用のクエリ言語 SPARQL のサポートも提供される。

TDB は Apache Jena のコンポーネントの 1 つである RDF ストレージであり、RDF を操作するための API や SPARQL エンジン、RDF 用データベースである。TDB には、各種コマンドラインツール、SPARQL サーバーなどで構成され、Jena の API から扱うことが出来る。RDF データに対する推論ルールを設定することも可能である。TDB は軽量でシンプルな RDF ストレージであるが、データの挿入及び検索のパフォーマンスが比較的高く、扱いやすい RDF ストレージである。

4. SIBM - RDF 形式で記述した避難場所情報ベンチマーク

4.1 SIBM の概要・特徴

SIBM が避難場所情報と避難作業の関係者情報を生成可能である。

SIBM は以下の 2 つの目標に基づいて設計されている：

- (1) 避難する際、実際に発生する事情を再現できるデータを生成すること。
- (2) 生成した情報を問い合わせするクエリセットを用意すること。

その以外にも、研究分野や使用目的に応じて、様々な場面で活用できることが考えられる。

SIBM では、避難場所情報を現実に近い情報源を生成する目的を持ち、以下の 2 つの特徴をもつ：

- ・ SIBM では、指定された避難場所数によるデータを生成することが可能であり、生成した情報が実際の避難場所情報や被災者間の関係などを再現できる。
- ・ SIBM が生成した情報に対する問い合わせクエリセットを用意し、LUBM のような RDF データ構造に対するクエリや、実際の情報を問い合わせするケースに対するクエリを作成する。

以下は、SIBM が用意したクエリセットとなり、そのような問い合わせ例に対するデータモデルを実装する。

(1) LUBM を参考したクエリ

LUBM が RDF の構成やデータストレージの性能の評価に対するクエリを作成した。ここは、LUBM が構成したクエリのいくつかを参考した避難場所の RDF データを問い合わせセットを作成する。

Query 1: LUBM のクエリ 1

```
SELECT ?X
WHERE {
  ?X rdf:type ub:GraduateStudent .
  ?X ub:takesCourse
    http://www.Department0.University0.edu/
    GraduateCourse0
}
```

Query 2: lubm_insp_q1 : 「sc1」学校で勉強している学生を検索するクエリである。検索クラスは人全体であり、推論情報やデータの階層を持たないクエリである。このクエリが LUBM のクエリ 1 を再現する。

```
SELECT ?x
WHERE {
  ?x sIBM:studyAt <http://lab.ene.im/SIBM/thing/school/sc1> .
}
```

Query 3: LUBM のクエリ 2

```
SELECT ?X, ?Y, ?Z
WHERE {
  ?X rdf:type ub:GraduateStudent .
  ?Y rdf:type ub:University .
  ?Z rdf:type ub:Department .
  ?X ub:memberOf ?Z .
  ?Z ub:subOrganizationOf ?Y .
  ?X ub:undergraduateDegreeFrom ?Y
}
```

Query 4: lubm_insp_q2 : 「小学校」で勉強している学生たちの情報及び避難している場所の情報を検索するクエリである。クエリ 1 より複雑な検索条件で検索するクエリである。このクエリが LUBM のクエリ 2 を再現する。

```
SELECT ?x ?y ?z
WHERE {
  ?x sIBM:stayAt ?y .
  ?x sIBM:studyAt ?z .
  ?z a "小学校"^^xsd:string .
}
```

Query 5: LUBM のクエリ 3

```
SELECT ?X
WHERE {
  ?X rdf:type ub:Publication .
  ?X ub:publicationAuthor
    http://www.Department0.University0.edu/
    AssistantProfessor0
}
```

Query 6: lubm_insp_q3 : 避難場所「S1」にいる人の情報を検索するクエリである。クエリ 1 と似たようなクエリであるが、避難場所にはさらに複雑なデータ構造を持ち、LUBM のクエリ 3 を再現する。

```
SELECT ?x
WHERE {
  ?x sIBM:stayAt
    <http://lab.ene.im/SIBM/thing/shelterpoint/S1> .
}
```

Query 7: LUBM のクエリ 4

```
SELECT ?X, ?Y1, ?Y2, ?Y3
WHERE {
  ?X rdf:type ub:Professor .
  ?X ub:worksFor <http://www.Department0.University0.edu> .
  ?X ub:name ?Y1 .
  ?X ub:emailAddress ?Y2 .
  ?X ub:telephone ?Y3
}
```

Query 8: lubm_insp_q4 : 人の個人情報に対するクエリである。人に名前、年齢、メールアドレスを問い合わせる。LUBM のクエリ 4 を再現する。

```
SELECT ?x ?name ?age ?mail
WHERE {
  ?x sIBM:profile ?y .
  ?y sIBM:firstName ?name .
  ?y sIBM:age ?age .
  ?y sIBM:email ?mail .
}
```

Query 9: LUBM のクエリ 6

```
SELECT ?X
WHERE {
  ?X rdf:type ub:Student
}
```

Query 10: lubm_insp_q6 : 避難場所の関係者のタイプに関するクエリである。「ボランティア」である人を検索する。LUBM のクエリ 6 を再現する。

```
SELECT ?x
WHERE {
  ?x rdf:type
    <http://lab.ene.im/SIBM/thing/usertype/volunteer> .
}
```

Query 11: LUBM のクエリ 7

```
SELECT ?X, ?Y
WHERE {
  ?X rdf:type ub:Student .
  ?Y rdf:type ub:Course .
  ?X ub:takesCourse ?Y .
  <http://www.Department0.University0.edu/
    AssociateProfessor0> ub:teacherOf ?Y
}
```

Query 12: lubm_insp_q7 : クエリ 6 の拡張であり、さらに複雑な検索クエリである。LUBM のクエリ 7 を再現する。

```
SELECT ?x ?y
WHERE {
  ?x rel:childOf ?y .
  ?y rdf:type
    <http://lab.ene.im/SIBM/thing/usertype/assistant> .
  ?y sIBM:stayAt
    <http://lab.ene.im/SIBM/thing/shelterpoint/S1>
}
```

(2) 実際の使用シナリオに対するクエリ

Query 13: select_label : ラベルが避難者の怪我の状態である。このクエリでは、避難者の状態を集約・分類するクエリである

```
SELECT ?label (count(?p) AS ?count)
WHERE {
  ?p sIBM:label ?label
} group by ?label order by ?label
```

Query 14: select_age : 年齢が 60 より高い人数を検索するクエリである。実際に、避難者の年齢・性別などの情報を検索することが多い。

```
SELECT (count(*) AS ?count)
WHERE {
  ?x sIBM:profile ?y .
  ?y sIBM:age ?z ;

  FILTER(?z > "60"^^xsd:integer)
}
```

Query 15: select_family : 避難者の中に、病状がよくない人の家族情報を検索するクエリである。

```
SELECT ?Principal ?Child ?Father
WHERE {
  ?x sIBM:profile [sIBM:firstName ?_Principal] .
  ?x sIBM:status <http://lab.ene.im/SIBM/thing/status/serious>
  .

  OPTIONAL {
    ?x sIBM:hasChild ?y .
    ?y sIBM:profile [sIBM:firstName ?_Child]
  }

  OPTIONAL {
    ?x sIBM:hasFather ?z .
    ?z sIBM:profile [sIBM:firstName ?_Father]
  }
}
```

Query 16: select_gender_ext : 避難者の性別で分類するクエリである。

```
SELECT ?gender (count(?p) AS ?count)
WHERE {
  ?p sIBM:profile ?y .
  ?y sIBM:gender ?gender ;
} group by ?gender order by ?gender
```

Query 17: select_storage : 避難場所の備蓄状況を確認するクエリである : 米の残量が 100 人以上がある場所を検索する .

```
SELECT ?storage ?p
WHERE {
  ?x sibm:address ?p .
  ?x sibm:storageRice ?storage .
  FILTER(?storage > 100)
} order by ?storage
```

Query 18: select_school : 小学生の家族の情報を検索するクエリである . 実際に、親との連絡などよくあるクエリだと考えている .

```
SELECT ?x
WHERE {
  ?x rel:parentOf ?y .
  ?x sibm:studyAt ?z .
  ?z a "小学校"^^xsd:string .
}
```

4.2 データ構成の設計

上記のクエリに対し、SIBM が 3 つの情報 : 避難場所情報、関係者の情報とその関連情報 (関係者の所属情報や避難場所の物品情報など) を対象とする . その 3 つの要素からなるデータセットが図 4 のように構成される .

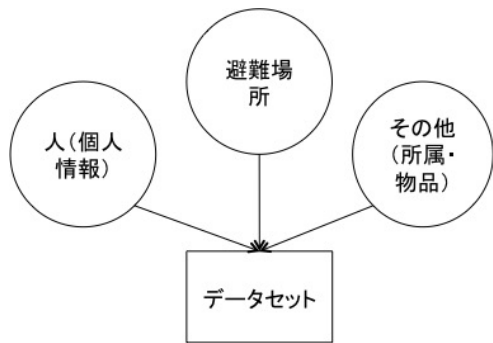


図 4: データセットの構成

避難場所情報が避難場所の詳細情報を表すものである . その構造は以下ようになる :

- 地上の座標、住所
- 収容人数
- 災害分類 (地震・津波・火山・洪水・その他・未確定)
- 施設の種類

関係者が 3 つのタイプある : アシスタント、避難者、ボランティア . 関係者全体に対する基本個人情報生成され、タイプ別に対する所属情報や怪我状態などの情報が追加される . 例えば、一般個人情報には、ユーザの氏名・年齢・住所・性別などの情報が持つ . ユーザタイプは避難者であるならば、怪我情報が追加され、また、ユーザの年齢によって会社や学校などに所属する情報が追加される .

避難場所情報全体には、避難場所の詳細情報と、その避難場所避難する・作業する人たちの情報が含まれる . 関係者の分配割合が適切に生成される (図 5) .

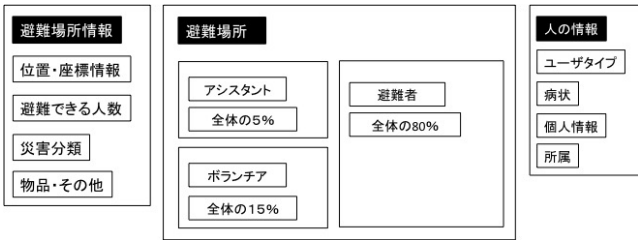


図 5: 避難場所全体の構成

実際、被災地には、一人以上の家族単位で避難することが考えられ、また、家族全員が違う避難場所避難することが考えられる . SIBM では、そのような関係を追加することが可能である (図 6) .

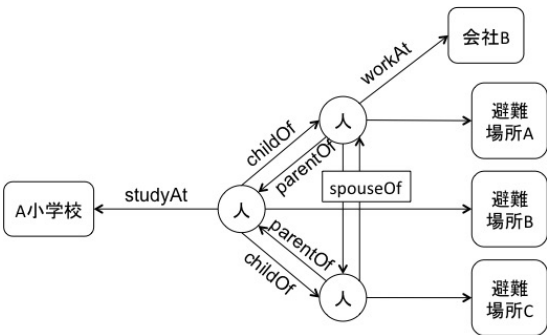


図 6: 避難場所に置ける情報構成

最後に、生成した避難場所情報を都道府県単位でまとめる (7) . 本稿では、平成 2 4 年度国土交通省による公開された全国の避難場所情報 (役 12.5 万箇所) を使用した .



図 7: 都道府県単位の避難場所構成

4.3 SIBM の実装

4.3.1 実装環境

SIBM が Apache Jena (3.3) を利用して、Java プログラミング言語で実装される .

4.3.2 データの生成

SIBM が現実に近い情報を生成するために、実際にあるリソースを使用することと、実際に近いデータモデルを生成する . 上記の実装を行うための手順や使用したリソースについて説明する .

- 避難場所情報

SIBM が国土交通省 (MLIT^(注3)) による公開した避難施設情報^(注4)を使用した (ただし、最新の情報ではない場合もあるこ

(注 3) : Ministry of Land, Infrastructure, Transport and Tourism
(注 4) : <http://nlftp.mlit.go.jp/ksj/index.html>

とから、災害時に避難施設を利用できることにならない)。公開した避難場所情報の詳細が(表3)になる。

表 3: 避難場所情報の詳細

データ項目	フォーマット	整備年度	内容
避難施設	点	平成 24 年度	位置、行政区域、名称、住所、施設の種類、収容人数、施設規模、災害分類

また、避難場所の備蓄や、避難場所への支援物品・薬品などの情報が避難場所の収容量や災害発生位置への距離による人工的に生成される。生成された情報を Jena API による RDF データ形式に出力する。避難場所の RDF データの詳細が表 4 に表示される。

表 4: 避難場所情報の RDF 述語の詳細

フィールド	RDF 述語	内容
位置	geo:geopoint	避難場所の位置
行政区域	sibm:administrativeAreaCode	都道府県コードと市区町村コードからなる、避難場所の所在する行政区を特定するためのコード
名称	sibm:name	避難場所の名称
住所	sibm:address	避難場所の住所
施設の種類	sibm:facilityType	避難場所の分類
収容人数	sibm:seatingCapacity	避難施設の形態ごとの収容可能人数
施設規模	sibm:facilityScale	避難施設の形態ごとの面積
災害分類	sibm:hazardClassification	当該施設が対象とする災害の分類
備蓄	sibm:storage	避難場所の備蓄情報、人分単位で表現する

・ 関係者・個人情報 SIBM では、実際の人間の個人情報ではないが、実際に近い情報を生成することができる。SIBM が表 5 のような個人情報を生成するツールを構成する。構成したツールを使用して個人情報を生成することができ、さらに複雑な情報(家族関係・所属情報など)を生成することも可能である。

ただし、生成する人間情報の男女の割合や年齢の配布が日本人口構造情報をにより行う。人は家族単位で生成され、家族の人数が「深さ(depth)」パラメーターによる決められる。例えば、深さ 0 の家族は一人の場合となるが、深さ最大値は 2 の場合、3 世代の家族になる。家族情報を生成する際、最小の子供の年齢から始め、日本人口の「出産平均年齢」の情報と「平均子供の数」の情報により、子供の数や親の年齢など家族に関する関連情報を生成する。

最後に、個人情報の中に、人の年齢、性別、タイプによる所属情報・怪我の状態などを生成する。

表 5: 個人情報の例

フィールド	値
firstName	NLLg
surname	Vt
gender	Female
birthday	1982-6-5
age	33
phone	+81xx-xxxx-xxxx
email	NLLg_Vt@Caa.com
workAt	株式会社 AAzX

SIBM が入力した避難場所数に対する避難場所情報や関係者情報を生成可能である。ただし、災害発生時、近くの避難場所から避難することが考えられるため、最初に SIBM が「災害発生中心」を適当に生成し、その中心から離れる避難場所を生成する。そしてまた、避難する際、家族の人がなるべく近い避難場所で避難するや作業することが考えられる。そのようなシナリオを実装するために、生成された家族の人数により、避難場所リストから特定な数だけの避難場所を使用し、家族全員を避難場所に適切に分配する。

SIBM では、避難場所数によるデータセットのスケールが決められる。入力した避難場所数に応じる人数を生成することで、任意なデータセットスケールを生成することが可能である。ただし、日本には 12.5 万程度の避難場所があるため、より多い避難場所を生成することは有りえないと判定する。

ここまで、避難場所情報・関係者情報及びそれらに関する関連情報を生成することができる。

4.3.3 SPARQL クエリセット

SIBM で生成された情報をクエリすることにより、RDF データ管理システムを評価することができる。避難場所情報の適用シナリオに対するクエリセットを構成する。

SIBM では、避難場所情報だけでなく、RDF による情報を記述したデータも扱う目的を持ち、LUBM を参考したいいくつかの RDF データ構造に対するクエリを用意する。その一方、実際の使用シナリオに対するクエリセットも用意する。クエリセットの詳細は 4.1 のように生成した。

5. 実験・考察

構成した SIBM を以下の実験環境(表 6)で実験を行い、データの生成時間やデータベースシステムとのアクセス時間についての考察を行う。

SIBM を用いて生成したデータが図 8 のよになる。

大数の避難場所情報を生成する場合、かかった時間やデータ量などの情報が表 7 のよになる。

図 9 でわかるように、避難場所数によるデータサイズの変動が確認できる。そこで、データサイズによる生成時間やデータベースへの挿入時間の変動を再現できることがわかる。

次に、生成したデータに対するクエリを行い、応答時間について考察する。なお、使用したクエリが 4.3.3 で説明する。本実験が Jena/TDB をデータストレージとして使用した。クエリの

表 6: 実験環境

Operation System	Windows 8.1 Pro 64 bits
CPU	Intel core i7 3770, 4 cores 8 thread @ 3.4Ghz (Boostable up to 3.9Ghz)
RAM	16GB
Storage	HDD
Java	1.7.0_71 (64 bits)
Apache Jena	2.12.1
SIBM	build 20150206

```

@prefix geo: <http://lab.ene.im/SIBM/property/geo#> .
@prefix sIBM: <http://lab.ene.im/SIBM/property#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://lab.ene.im/SIBM/thing/shelterpoint/S5>
  sIBM:address "北海道夕張市南部東町番地
    76"^^xsd:string ;
  sIBM:administrativeAreaCode "1209"^^xsd:int ;
  sIBM:facilityScale "-1"^^xsd:int ;
  sIBM:facilityType "収容避難所"^^xsd:string ;
  sIBM:hazardClassification
    [ sIBM:earthquakeHazard false ;
      sIBM:notSpecified false ;
    ] ;
  sIBM:name "南部コミュニティセンター"^^xsd:string ;
  sIBM:seatingCapacity "400"^^xsd:int ;
  geo:geopoint
    [ geo:latitude "43.020741"^^xsd:double ;
      geo:longitude "142.08"^^xsd:double
    ] .

<http://lab.ene.im/SIBM/person/cb6cb515-8415-4466-b24e-741
d1e841499>
  a <http://lab.ene.im/SIBM/thing/usertype/evacuee> ;
  sIBM:label <http://lab.ene.im/SIBM/thing/label/Injured3> ;
  sIBM:profile
    [ sIBM:age "19"^^xsd:int ;
      sIBM:birthday "1996-8-13"^^xsd:string ;
      sIBM:email "KgoG_DEr@wSdl.com"^^xsd:string ;
      sIBM:firstName "KgoG"^^xsd:string ;
      ...
    ] ;
  sIBM:stayAt <http://lab.ene.im/SIBM/thing/shelterpoint/S5>
    > ;
  sIBM:studyAt <http://lab.ene.im/SIBM/thing/school/
    sc1043> ;
  rel:childOf <http://lab.ene.im/SIBM/person/bff83765-
    b629-480e-a362-6f6c3f2fca6> ,
    <http://lab.ene.im/SIBM/person/97f7ea59-3
    f8e-4892-8957-08d93f61522b> .

<http://lab.ene.im/SIBM/thing/school/sc1043>
  a "大学"^^xsd:string ;
  sIBM:address "緑泉町 5-14"^^xsd:string ;
  sIBM:administrativeAreaCode "1216"^^xsd:int ;
  sIBM:name "星槎大学"^^xsd:string ;
  geo:geopoint
    [ geo:latitude "43.459416"^^xsd:double ;
      geo:longitude "142.166334"^^xsd:double
    ] .

```

図 8: 実装例

表 7: 避難場所数に対する生成時間

場所数	生成 (ms)	関係者数	トリプル数	挿入 (ms)	サイズ (MB)
5	3373	1455	20409	3807	203.87
10	4779	3018	42304	3669	205.57
20	6783	6082	85482	5588	223.91
50	7603	12645	177991	8312	296.59
100	6591	24751	348443	12662	379.90
200	10363	53719	755895	22522	625.34
350	24708	112888	1584015	43468	1127.90
500	30228	152903	2148284	55275	1456.53
700	42229	220201	3091954	77798	2004.87

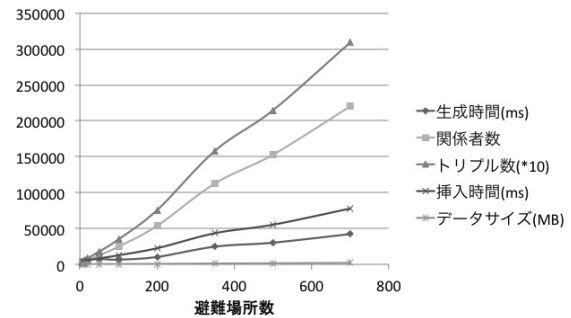


図 9: 生成・挿入時間とデータサイズ

応答時間は以下の表 8、表 9 及び図 10 のようになる。

表 8: クエリ時間 (ms) (LUBM を参考したクエリセット) (「li」は「lubm_insp」の略となる)

場所数	li_q1	li_q2	li_q3	li_q4	li_q6	li_q7
5	122	9	11	157	4	7
10	203	9	12	240	4	6
20	99	8	11	289	5	4
50	96	9	14	375	6	4
100	94	8	11	494	8	4
200	97	9	11	799	13	4
350	98	9	11	2157	91	4
500	96	9	10	4576	32	4
700	97	9	12	7779	82	4

表 9: クエリ時間 (ms) (使用シナリオに応じたクエリセット) (「select_」を略すること)

場所数	label	age	family	gendet_ext	storage	school
5	41	63	9	39	23	3
10	46	73	9	39	4	3
20	58	89	9	67	10	3
50	69	155	13	114	5	3
100	102	183	8	109	6	3
200	123	281	8	195	9	3
350	154	492	8	412	18	3
500	174	706	9	531	25	3
700	351	1102	8	1192	40	3

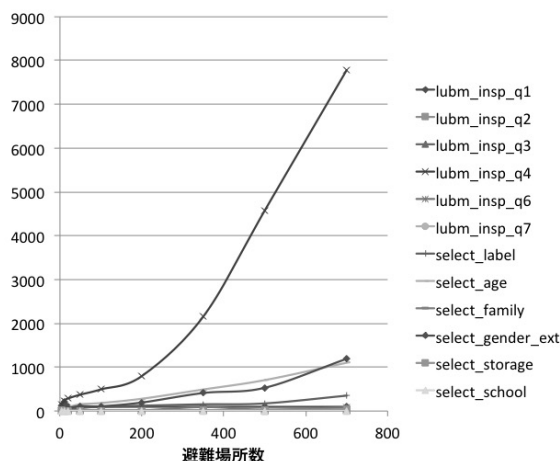


図 10: クエリ時間

ここでは、避難場所数による時間がかかるクエリもあり、避難場所数に対する時間変動のないクエリもあるとわかる。

時間がかかるクエリ (lubm_insp_q4、select_label、select_age、select_gender_ext) では、クエリするデータの構造が複雑であることがあり、または複雑な検索情報の条件でもある。

その一方、他のクエリでは、階層構成を持たないデータ (学校の情報など) や、検索条件が単純であるクエリの場合に対して、時間変動がほとんどない、または少ないと言える。

そので、SIBM では、多種のクエリで、様々な実験シナリオを再現できることが確認できる。

6. 結論、今後の課題

6.1 結 論

近年、RDF 形式で災害情報などを含め様々な情報を共有することが増えてきて、そのような情報を管理する適切なシステムが求められる。

本研究は、現実に近い避難場所情報を生成可能な SIBM というベンチマークを提案し、それらの情報を問い合わせる現実に近いクエリセットを用意することで、そのようなシステムを評価することを可能にした。

6.2 今後の課題

構成したベンチマークに対して、いくつかの解決必要な問題が考えられる。

避難場所における事情が時間とともに変動することが事実である。本研究は、そのような変動を再現できず、固定な情報へのアクセスしかできていない。そのため、情報の歴史や、時間に対する変動な数値に応じる様々な変動に対するベンチマークを構成することを解決すべき課題であると考えている。

また、本稿の実験では Jena/TDB だけを使用した。多種の RDF データストレージで実験を行い、それらのシステムの性能を評価、比較することが今後の課題と考えている。

文 献

- [1] Jens Ortmann, Minu Limbu, Dong Wang, and Tomi Kauppinen “Crowdsourcing Linked Open Data for Disaster Management”

- [2] Yuanbo Guo, Zhengxiang Pan, and Jeff Heflin “LUBM: A Benchmark for OWL Knowledge Base Systems”
- [3] 児玉 快、横田 治夫、“データやユーザの効率的な追加・削除が可能な秘匿情報アクセス手法”、第 7 回データ工学と情報マネジメントに関するフォーラム、2015
- [4] Vu Tuan Dat, 横田 治夫、“MapReduce による大規模な RDF データ復号化手法の評価”、第 7 回データ工学と情報マネジメントに関するフォーラム、2015
- [5] M. Fujiwara, “KSJ - 国土数値情報の変換プログラム”, <https://github.com/ma38su/ksj.git>
- [6] Data of Japan project, <https://github.com/dataofjapan/land.git>
- [7] Apache Jena - A Semantic Web Framework for Java, “<https://jena.apache.org>”
- [8] 東京都福祉保健局 - “避難所管理運営の指針 (区市町村向け)”
- [9] 復興庁 - “避難所生活者・避難所の推移 (東日本大震災、阪神・淡路大震災及び中越地震の比較)”