

東京工業大学工学部

学士論文

SIBM - RDF形式に記述した避難場所情報
ベンチマークツール

指導教員 横田 治夫 教授
荒堀 喜貴 助教

平成27年2月

提出者

学科 情報工学科
学籍番号 11_27820
氏名 NGUYEN Hoai Nam

指導教員認定印		
学科長認定印		

SIBM - RDF 形式に記述した避難場所情報 ベンチマークツール

指導教員 横田 治夫 教授
荒堀 喜貴 助教

情報工学科

11_27820 NGUYEN Hoai Nam

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、被災地における被災者-支援者間のデータ共有方式の妥当性の評価を可能にすることである。具体的には、被災者の怪我の程度などの個人情報医療従事者にのみ開示し、他の被災者には個人情報へのアクセスを許可しないなどのデータ共有方式の性能や秘匿情報保護能力を SIBM によって計測できるようになることを目指す。

目 次

第1章 序論	1
第2章 避難場所に置ける情報の管理・使用例	2
2.1 避難場所情報の検索	2
2.2 避難場所における情報集約	3
第3章 関連研究	4
3.1 Crowdsourcing Linked Open Data for Disaster Management . . .	4
3.2 LUBM - Lehigh University Benchmark	4
3.3 その他	5
第4章 前提知識	6
4.1 RDF について	6
4.2 SPARQL クエリ言語	6
4.3 Apache Jena と TDB	8
第5章 SIBM - 避難場所情報ベンチマークツール	9
5.1 SIBM の概要	9
5.2 SIBM の特徴	9
5.3 データ構成の設計	10
5.4 SIBM の実装	11
第6章 実験・考察	16
第7章 結論・今後の課題	20
7.1 結論	20
7.2 今後の課題	20
謝辞	21
参考文献	21
付録 A クエリセット	23
A.1 LUBMを参考したクエリ	23

A.2 実際の使用シナリオに対するクエリ	24
--------------------------------	----

第1章

序論

全世界には、毎年災害が沢山発生する。大規模災害の2011年の大震災の他、台風による災害事故や土砂災害などによる破壊が多いとわかった。災害が発生するとき、避難することは重要である。避難する際、避難場所情報の管理や、避難する作業に関係のある人々の情報が多量に増加することがある。それらの情報を管理・アクセスすることが重要な作業である。現実では、避難することに関する情報の中に、避難場所情報、避難する人の情報、避難場所で作業する人（アシスタント・ボランティアなど）の情報があり、それらの情報を管理・アクセスすることが考えられる。災害が発生するとき、それらの情報に対してすぐに対応できることが重要である。そのため、災害対策作業に対して、現実に近い情報を使用しながら実験することは実際の要求である。

一方、近年には、様々な新たなデータ記述法が開発されている。その中に、RDF形式で記述されたデータが増えている。RDF形式で記述された避難・災害情報を公開・管理することが考えられている[1]。RDF形式で記述したデータが、情報へのアクセス範囲をさらに細かく設計できることがわかる。避難作業情報の中に、そのようなアクセス制限モデルが必要であり、共有範囲を限定すべき情報を暗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、避難者にみせられない情報や個人情報の秘密性に応じるアクセス範囲の制限が必要と考えられる[3]。また、災害が発生するとき、情報がだんだん増加することがわかる。多量データを暗号化することも考えられる。そこで、効率的に暗号化する手順が必要である[4]。

上記の問題に対して、RDF形式で記述した避難する作業に関するデータセットを使用して様々な使用シナリオで実験する要求がある。本研究では、RDF形式で記述された避難場所情報と避難作業情報を生成するベンチマークツール - SIBM - を実装する。そして、実装したベンチマークツールの構成、実験環境と実用例について考察する。

第2章

避難場所に置ける情報の管理・使用例

避難所は、区市町村があらかじめ指定している避難施設で、災害時等に区市町村長が開設・管理・運営し、避難者に安全・安心の場として提供することを目的としている。

大規模な災害発生時に、体育館等の施設を避難所として開放し、大量の避難者を受け入れる上では、避難者情報を管理するシステムが求められる。避難場所情報に対する様々な使用目的があり、その中によく考えられるのは情報の検索や情報の集約である。

2.1 避難場所情報の検索

災害発生時、避難者情報や避難場所で作業する人の情報などの情報を検索することがある。避難場所情報の検索についていくつかの例がある。

1. 災害発生時、その周辺の近くに避難することが考えられ、適切に避難できるために、位置情報による避難場所を検索することがある。
2. 避難する時、大量の人々が避難先へ水平移動をすると、避難先の収容量を超えてしまったり、避難路における混雑が発生したりすることがある。そこで、避難場所の収容量やその他の関連情報を検索することがある。
3. 避難する際、個人情報も含め、避難者の状況や避難する場所などの情報を検索することが多い。
4. 災害発生のため、ある家族の人たちが違う避難場所で避難することになり、家族関係による検索の適用で、それらの家族の情報を検索場合がある。または、情報のアクセスすることに対して、他人がアクセス許可を持たない情報でも、親戚宅・知人宅等がアクセスできると安心であるため、そのような関係を考慮しながらアクセス許可を制御することがある。

5. 避難所においては、被災者の視点に立って、高齢者、障害者、乳幼児、妊産婦等の災害時要援護者や、女性に対する配慮が必要であることから、年齢・性別・健康状況による検索場合がある。

2.2 避難場所における情報集約

情報を検索するだけでなく、検索した情報の集約、分析などの作業が考えられる。例えば：

1. 災害発生時、世界の国々からの国際的なサポートがある。薬品・食品などを適切に避難場所へ配ることが大切であり、避難者の状況・病類による物品・薬品の分配することが考えられる。そのような作業は避難場所全体の情報を検索した数値による行うことになる。
2. また、避難場所に分配された食品などの物品に対して、なるべく避難者ありうは役員に適切に分配する必要がある。例えば、高齢者のための食品では、使用量や種類が大人に対して違うことや、子供の場合も同様に考えられるだろう。その時、避難者や役員の年齢・性別情報により、物品の使用を決まることがある。
3. 避難するさい、物品や食材などの用品が時間とともに増減である。その対応に当たって、避難場所情報の管理システムによる、使用歴史や統計などを集約することもある。

本研究では、避難場所に置ける情報を RDF 形式で記述したデータの管理システムを位置づけすることで、上記のような実際の使用シナリオを考慮しながら、避難場所情報管理システムの設計や運営に置ける必要な評価をするためのベンチマークを構成する。そして、上記の例に対するいくつかのクエリを実験することにより、構成したベンチマークを考察する。

第3章

関連研究

3.1 Crowdsourcing Linked Open Data for Disaster Management

近年、災害情報の管理に当たって、災害情報のデータ工学の重要な問題とみなして、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るとわかった。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。

3.2 LUBM - Lehigh University Benchmark

LUBM (Lehigh University Benchmark) が RDF/OWL データのベンチマークである。LUBM では、大学情報 (大学・学部・授業) 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た。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る。[2] で Guo らが LUBM を用いることで、RDF/OWL ストレージの評価を行った。それより、実際にあるデータを扱う時、適切なシステムを設計することについて論じた。

3.3 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。

Jens らの研究 [1] を始め、RDF の適用可能性により、RDF 形式で記述した災害情報・避難場所情報管理システムの必要性がわかり、現実に近い避難場所情報を用いることで RDF 情報管理システムを評価することは本研究のモチベーションである。そうするために、LUBM のようなベンチマークによるデータ生成ツールを使用することで実現できると考える。

しかしながら、LUBM で生成したデータの中に、個人情報や、関係者間の関係が不足だと考え、第 2 章に記述した実際の使用シナリオに適用可能性が低い。さらに、避難場所情報が特殊な情報であり、LUBM が生成したデータのままで適用できないことがわかった。

本研究では、LUBM のデータ生成モデルを参考することで、RDF 形式で記述した避難場所情報を含め、関係者間の関係や実際にある使用シナリオを再現できるベンチマークを構成する。また、生成したデータのもとで、従来の災害対策のための情報管理システムを評価するために、現実に近い情報とその情報にたいするクエリを構成できることを目指している。

第 5 章から、本研究で提案した SIBM - Shelter Information Benchmark について説明する。

第4章

前提知識

4.1 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。^{*}

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の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ーとプロパティー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、そのプロパティーと値として、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 4.1 の A) あるいは RDF グラフ (トリプルの集合) (図 4.1 の B) で表現できる。

4.2 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

^{*}<http://www.asahi-net.or.jp/~ax2s-kmttn/internet/rdf/rdf-primer>

Fig. 4.1: RDF 構成の例

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。[†]

以下は、SPARQL の例である。図 4.3 が RDF グラフ 4.2 のためのクエリであり、そのクエリ結果が表 4.1 となる。

```
1  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
2
3  _:a foaf:name Johnny Lee Outlaw .
4  _:a foaf:mbox <mailto:jlow@example.com> .
5  _:b foaf:name Peter Goodguy .
6  _:b foaf:mbox <mailto:peter@example.org> .
7  _:c foaf:mbox <mailto:carol@example.org> .
```

Fig. 4.2: RDF サンプル

[†]<http://www.asahi-net.or.jp/~ax2s-kmtn/internet/rdf/REC-sparql11-query-20130321>

```

1 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
2 SELECT ? name ? mbox
3 WHERE {
4     ? x foaf:name ? name .
5     ? x foaf:mbox ? mbox
6 }

```

Fig. 4.3: SPARQL サンプル

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

TABLE 4.1: クエリ結果

4.3 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。

第5章

SIBM - 避難場所情報ベンチマークツール

5.1 SIBMの概要

SIBM (Shelter Information Benchmark) は避難場所情報と避難作業の関係者に関する情報を中心するデータを生成するツールである。

SIBMは以下の2つの目標に基づいて設計されている：

1. 第2章に記述した使用シナリオから始め、避難する際、実際に発生する事象を再現できるデータを生成すること。
2. 拡張性を持つこと：任意の規模を持つデータセットやそのデータの複雑度を選択できることが実験に対する重要な要素であると考えている。

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。

5.2 SIBMの特徴

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3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る。また、実際の使用例を再現するクエリセットによるベンチマークを行うことで、データベースシステムを評価することができる。
- SIBMによる生成した情報の中に、実際の避難場所を使用することや、日本人口構成と実際にある人間の関係による関係者情報を生成することで、実際に近い情報源を生成することが可能である。

- 避難場所情報・人個人情報・その他の情報（人の就活先・配属情報や外部からのサポートなど）のマッピングによる、実際の情報を再現できる。

5.3 データ構成の設計

SIBM が 3 つの情報：避難場所情報、関係者の情報とその 2 つの情報と関係がある他の情報（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 5.1 のように構成される。

避難場所情報が避難場所の詳細情報を表すものである。その構造は以下のようになる（図 5.1）：

- 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 つの種類がある：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、タイプ別に対する情報が追加される。例えば、ユーザ全体の情報には、ユーザの氏名・年齢・住所・性別などの情報が持つ。情報共有範囲を管理するために、各タイプそれぞれに対するアクセス可能範囲の情報が追加される。

関係者の状況・タイプ・年齢などの情報による、所属情報などを生成される。避難場所にたいしては、収容人数情報や面積情報に対する物品情報を追加される。

避難場所情報全体には、避難場所の詳細情報と、その避難場所で避難する・作業する人たちの情報が含まれる。避難する人や避難場所で作業する人の割合が適切に生成される（図 5.2）。また、現実には、災害が発生するとき、一人以上の家族単位で避難することが考えられる。

また、家族全員が違う避難場所で避難することが考えられ、SIBM では、そのような関連を追加することが可能である（図 5.3）。

最後に、現実のように避難場所が都道府県単位で管理することを表す。本研究では、平成 24 年度国土交通省による公開された全国 47 都道府県の避難場所情報（約 12.5 万箇所）を使用した。

Fig. 5.1: データセットの構成

Fig. 5.2: 避難場所全体の構成

5.4 SIBMの実装

5.4.1 実装環境

SIBMがApache Jena（第4章）を利用して、Javaプログラミング言語で実装されている。

5.4.2 データの生成

SIBMでは、現実に近い情報を生成する目標を持ち、その目標を実装するために、実際にある情報元を使用することと、実際に近い情報を生成するツール

Fig. 5.3: 避難場所に置ける情報構成

Fig. 5.4: 都道府県単位の避難場所構成

を使用した。上記の実装を行うための手順や関連するリソースについて以下のように説明する。

- 避難場所情報については、SIBM が**国土交通省**（MLIT*）による提供した避難施設情報[†]を使用した（ただし、最新の情報ではない場合もあることから、災害時に避難施設を利用できないことにならない）。提供した避難場所情報の詳細が（表 5.1）になる。

*Ministry of Land, Infrastructure, Transport and Tourism

[†]<http://nlftp.mlit.go.jp/ksj/index.html>

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

TABLE 5.1: 避難場所情報の詳細

また、避難場所の備蓄や、避難場所へのサポート物品・薬品などの情報が避難場所の収容人数や災害発生位置への距離による人口的に生成される。最後に、生成された情報を Jena による RDF データ形式に出力する。避難場所情報における RDF データの詳細が表 5.2 に表示される。

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

TABLE 5.2: 避難場所情報の RDF 述語の詳細

- 人間の情報については、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 5.3 のような人間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人間の個人情報を生成することができ、さらに複雑な情報（家族関係など）を生成することも可能である。

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

TABLE 5.3: 個人情報の例

ただし、生成された人中に、男女の割合や年齢の配布が日本人口構造情報による生成される。人は家族単位で生成され、家族の「深さ (depth)」パラメーターによる家族の人数が決められる。

例えば、深さ 0 の家族は一人の場合にあり、3 世代の家族まで検討しながら、深さ最大値は 2 に設定する。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他の情報を生成する。

最終的に、個人情報の中に、人の年齢、性別、タイプによる所属情報などを生成する。

生成された避難場所情報は基本的、ある都道府県にある避難場所のリストになり、災害発生位置から離れるような形になる。実際では、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、家族全員を特定された避難場所に適切に分配する。

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。

5.4.3 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。ここで、第 2 章に説明した避難場所情報の適用シナリオにたい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、第2章にある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付録Aに示す。

第6章

実験・考察

構成したSIBMを以下の実験環境（表6.1）で実装し、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考察を行う。

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	v0.9 build 20150129

TABLE 6.1: 実験環境

SIBMを用いて生成したデータが図6.1のような形になる。

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6.2のようになる。

図6.2でわかるように、避難場所数が生成時間や人数・トリプル数と比例関係が確認できる。

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Aに説明する。また、Jena/TDBをデータストレージとして使用した。実行時間をグラフ化した物は以下の図6.3のようになる。

SIBMでは、特定のパラメーターに対する避難場所情報のRDFデータセットを生成することができ、現実に近い情報の元で実験を行うことが可能である。このセッションでは、SIBMで生成したデータセットとそのもとのパラメー

```

1  @prefix geo: <http://lab.ene.im/SIBM/property/geo#> .
2  @prefix sIBM: <http://lab.ene.im/SIBM/property#> .
3  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
4
5  <http://lab.ene.im/SIBM/thing/shelterpoint/S5>
6      sIBM:address "北海道夕張市南部東町番地 76"^^xsd:string ;
7      sIBM:administrativeAreaCode "1209"^^xsd:int ;
8      sIBM:facilityScale "-1"^^xsd:int ;
9      sIBM:facilityType "収容避難所"^^xsd:string ;
10     sIBM:hazardClassification [ sIBM:earthquakeHazard false ;
11                                 sIBM:notSpecified false ;
12                                 ...
13                                 ] ;
14     sIBM:name "南部コミュニティセンター"^^xsd:string ;
15     sIBM:seatingCapacity "400"^^xsd:int ;
16     geo:geopoint [ geo:latitude "43.020741"^^xsd:double ;
17                   geo:longitude "142.08"^^xsd:double
18                   ] .
19 <http://lab.ene.im/SIBM/person/cb6cb515-8415-4466-b24e-741d1e841499>
20     a <http://lab.ene.im/SIBM/thing/usertype/evacuee> ;
21     sIBM:label <http://lab.ene.im/SIBM/thing/label/Injured3> ;
22     sIBM:profile [ sIBM:age "19"^^xsd:int ;
23                  sIBM:birthday "1996-8-13"^^xsd:string ;
24                  sIBM:email "KgoG_DEr@wSdl.com"^^xsd:string ;
25                  sIBM:firstName "KgoG"^^xsd:string ;
26                  ...
27                  ] ;
28     sIBM:stayAt <http://lab.ene.im/SIBM/thing/shelterpoint/S5> ;
29     sIBM:studyAt <http://lab.ene.im/SIBM/thing/school/sc1043> ;
30     rel:childOf <http://lab.ene.im/SIBM/person/bff83765-b629-480e-a362-6f6c3f2fca6> ,
31                <http://lab.ene.im/SIBM/person/97f7ea59-3f8e-4892-8957-08d93f61522b> .
32 <http://lab.ene.im/SIBM/thing/school/sc1043>
33     a "大学"^^xsd:string ;
34     sIBM:address "緑泉町 5-14"^^xsd:string ;
35     sIBM:administrativeAreaCode "1216"^^xsd:int ;
36     sIBM:name "星槎大学"^^xsd:string ;
37     geo:geopoint [ geo:latitude "43.459416"^^xsd:double ;
38                   geo:longitude "142.166334"^^xsd:double
39                   ] .

```

Fig. 6.1: 実装例

ターとの反映性について考察する。また、生成した情報の実用性について考察する。

避難場所数	生成時間 (ms)	関係者数	トリプル数	挿入時間 (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

TABLE 6.2: 避難場所数に対する生成時間

Fig. 6.2: 生成・挿入時間とデータサイズ

SIBM バージョン 1 で対応した入力値は避難場所の数である。SIBM が生成できる最大の数は12.5万となる。ここでは、生成したデータを Jena/TDB データベースに挿入する時間と挿入後のデータサイズ・トリプル数について考察を行う。入力した避難場所に対する挿入時間とデータ量が表 6.2 になる。避難場所の関係者数を避難場所の規模による計算される。その数によって人間情報を生成する。表 6.2 で示すように、避難場所の数に応じた関係者の数がわかる。

避難場所数とそれに対するデータ生成・挿入時間の関係が図 6.2 に示す。生成されたデータ量（関係者数、トリプル数、データサイズ）が避難場所数に対

Fig. 6.3: クエリ時間

して線形関係があることが考えられる。だが、トリプル数/避難場所数の関係線が曲線になる。その理由は、特定な避難場所数に対して、生成しようとする関係者の数は定数ではなく、ある範囲内でランダムな数値で生成する。関係者情報に対するトリプル数と避難場所情報に対するトリプル数の総数では、ランダム性を持つため、直線な関係にならないからである。

また、Jena/TDB データベースでは、挿入するデータに対するメタデータやインデクシングすることによる発生したデータが多い。そのため、少量データに対してもデータベースサイズが 192.0MB になることがわかる*。

図 6.2 でわかるように、実際に発生する災害やそれに対する避難情報の増加が問題である。例えば、北海道には約 9000 避難場所がある。実際に全てを使用する要求がでると、多量な情報が発生することが考えられる。そして、災害が短時間で発生することが多いため、その短時間で多量なデータへの対策が事実な問題だとわかる。SIBM で生成するデータを使用し、その問題を解決する候補を実験することができる。

*<https://jena.apache.org/documentation/tdb/architecture.html>

第7章

結論・今後の課題

7.1 結論

7.2 今後の課題

謝辞

本研究を遂行し学位論文をまとめるにあたり、テーマの決定、研究の考え方、問題設計、まとめ方など全てにおいて、長期にわたって厳しくも熱意のあるご指導、ご鞭撻していただいた、横田治夫先生および荒堀喜貴先生に厚く御礼申し上げます。

参考文献

- [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”
- [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”
- [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>”

付録 A

クエリセット

A.1 LUBMを参考したクエリ

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
3
4 SELECT ? X
5 WHERE {
6   ? X rdf:type ub:GraduateStudent .
7   ? X ub:takesCourse
8       http:// www. Department0. University0. edu/ GraduateCourse0
9 }
```

Fig. A.1: LUBM のクエリ 1

```
1 PREFIX sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? x
4 WHERE {
5   ? x sibm:studyAt <http://lab.ene.im/SIBM/thing/school/sc1> .
6 }
```

Fig. A.2: 「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。このクエリが図A.1のLUBMのクエリ1を再現する。

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
3
4 SELECT ? X, ? Y, ? Z
5 WHERE {
6   ? X rdf:type ub:GraduateStudent .
7   ? Y rdf:type ub:University .
8   ? Z rdf:type ub:Department .
9   ? X ub:memberOf ? Z .
10  ? Z ub:subOrganizationOf ? Y .
11  ? X ub:undergraduateDegreeFrom ? Y
12 }

```

Fig. A.3: LUBM のクエリ 2

```

1 PREFIX sibm: <http://lab.ene.im/SIBM/property#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
4
5 SELECT ? x ? y ? z
6 WHERE {
7   ? x sibm:stayAt ? y .
8   ? x sibm:studyAt ? z .
9   ? z a "小学校"^^ xsd:string .
10 }

```

Fig. A.4: 「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。このクエリが図 A.3 の LUBM のクエリ 2 を再現する。

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
3
4 SELECT ? X
5 WHERE {
6   ? X rdf:type ub:Publication .
7   ? X ub:publicationAuthor
8       http://www.Department0.University0.edu/AssistantProfessor0
9 }

```

Fig. A.5: LUBM のクエリ 3

A.2 実際の使用シナリオに対するクエリ

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? x
4 WHERE {
5     ? x  sibm:stayAt <http://lab.ene.im/SIBM/thing/shelterpoint/S1> .
6 }

```

Fig. A.6: 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、図 A.5 の LUBM のクエリ 3 を再現する。

```

1 PREFIX  rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX  ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
3
4 SELECT ? X, ? Y1, ? Y2, ? Y3
5 WHERE {
6     ? X  rdf:type  ub:Professor .
7     ? X  ub:worksFor <http://www.Department0.University0.edu> .
8     ? X  ub:name ? Y1 .
9     ? X  ub:emailAddress ? Y2 .
10    ? X  ub:telephone ? Y3
11 }

```

Fig. A.7: LUBM のクエリ 4

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? x ? name ? age ? mail
4 WHERE {
5     ? x  sibm:profile ? y .
6     ? y  sibm:firstName ? name .
7     ? y  sibm:age ? age .
8     ? y  sibm:email ? mail .
9 }

```

Fig. A.8: 人の個人情報に対するクエリである。図 A.7 の LUBM のクエリ 4 を再現する。

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
3
4 SELECT ? X
5 WHERE {
6   ? X rdf:type ub:Student
7 }

```

Fig. A.9: LUBM のクエリ 6

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2
3 SELECT ? x
4 WHERE {
5   ? x rdf:type <http://lab.ene.im/SIBM/thing/usertype/volunteer> .
6 }

```

Fig. A.10: 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。図 A.9 の LUBM のクエリ 6 を再現する。

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX ub: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
3
4 SELECT ? X, ? Y
5 WHERE {
6   ? X rdf:type ub:Student .
7   ? Y rdf:type ub:Course .
8   ? X ub:takesCourse ? Y .
9   <http://www.Department0.University0.edu/AssociateProfessor0>
      ub:teacherOf ? Y
10 }

```

Fig. A.11: LUBM のクエリ 7

```

1 PREFIX sibm: <http://lab.ene.im/SIBM/property#>
2 PREFIX rel: <http://purl.org/vocab/relationship/>
3
4 SELECT ? x ? y
5 WHERE {
6   ? x rel:childOf ? y .
7   ? y rdf:type <http://lab.ene.im/SIBM/thing/usertype/assistant> .
8   ? y sibm:stayAt <http://lab.ene.im/SIBM/thing/shelterpoint/S1>
9 }

```

Fig. A.12: クエリ 6 の拡張であり、さらに複雑な検索クエリである。図 A.11 の LUBM のクエリ 7 を再現する。

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? label ( count(? p) AS ? count)
4
5 WHERE {
6   ? p  sibm:label ? label
7 } group by ? label order by ? label

```

Fig. A.13: 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

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2 PREFIX  xsd: <http://www.w3.org/2001/XMLSchema#>
3
4 SELECT ( count(*) AS ? count)
5 WHERE {
6   ? x  sibm:profile ? y .
7   ? y  sibm:age ? z ;
8
9   FILTER(? z >  " 60 "^^ xsd:integer)
10 }

```

Fig. A.14: 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。

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? Principal ? Child ? Father
4 WHERE {
5   ? x  sibm:profile [ sibm:firstName ? __Principal] .
6   ? x  sibm:status <http://lab.ene.im/SIBM/thing/status/serious> .
7
8   OPTIONAL {
9     ? x  sibm:hasChild ? y .
10    ? y  sibm:profile [ sibm:firstName ? __Child]
11  }
12
13  OPTIONAL {
14    ? x  sibm:hasFather ? z .
15    ? z  sibm:profile [ sibm:firstName ? __Father]
16  }
17 }

```

Fig. A.15: 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。

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? gender ( count(? p) AS ? count)
4 WHERE {
5     ? p  sibm:profile ? y .
6     ? y  sibm:gender ? gender ;
7 } group by ? gender order by ? gender

```

Fig. A.16: 避難者を性別で分類するクエリである。

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2
3 SELECT ? storage ? p
4 WHERE {
5     ? x  sibm:address ? p .
6     ? x  sibm:storageRice ? storage .
7     FILTER(? storage > 100)
8 } order by ? storage

```

Fig. A.17: 避難場所の備蓄状況を確認するクエリである。

```

1 PREFIX  sibm: <http://lab.ene.im/SIBM/property#>
2 PREFIX  xsd: <http://www.w3.org/2001/XMLSchema#>
3 PREFIX  rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4
5 SELECT ? x
6 WHERE {
7     ? x  rel:parentOf ? y .
8     ? x  sibm:studyAt ? z .
9     ? z  a  "小学校"^^ xsd:string .
10 }

```

Fig. A.18: 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。