

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 准教授

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 准教授
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 災害発生時に置ける避難場所情報の使用	2
2.1 避難場所の概要と役割	2
2.2 避難場所情報の検索	2
2.3 数値演算・物の分配	2
第3章 前提知識	3
3.1 RDF フレームワーク	3
第4章 関連研究	4
4.1 Crowdsourcing Linked Open Data for Disaster Management	4
4.2 LUBM - Lehigh University Benchmark	5
4.3 その他	5
第5章 SIBM - 避難場所情報ベンチマークツール	6
5.1 SIBM のコンセプト	6
5.2 データ構成の設計	6
5.3 データクラスの定義	8
5.4 情報リソース	10
第6章 実験・考察	12
6.1 実装例	12
6.2 考察	13
第7章 結論	16
参考文献	16
付録A クエリセット	18

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．本稿では、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．

災害発生時に置ける避難場所情報の使用

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。

結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1.
結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結
果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-1. 結果-
1. 結果-1. 結果-1. 結果-1. 結果-1.

結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．
結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結
果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-2．結果-
2．結果-2．結果-2．結果-2．結果-2．

第3章

前提知識

3.1 RDF フレームワーク

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 形式で記述された情報がトリプル (図 3.1 の A) あるいは RDF グラフ (トリプルの集合) (図 3.1 の B) で表現できる。

3.1.1 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。

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、各種コマンドラインツール、SPARQL サーバーなどで構成される。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。

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

Fig. 3.1: RDF の例

第4章

関連研究

4.1 Crowdsourcing Linked Open Data for Disaster Management

[1]では、近年、災害情報の管理がデータ工学の重要な問題とみなして、災害データ蓄積が巨大な革命であるとわかった．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．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わかった．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 Ortmann らが論じた．この研究で、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論じた．

4.2 LUBM - Lehigh University Benchmark

RDF/OWL データの実験ベンチマークについては Guo らの研究があり [2]、LUBM というベンチマークが開発されていた．LUBM では、大学情報やその関係者（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．しかし、現実世界のデータとの違いがあると考えられる．例えば、大学の関係者の種類別による大学の情報へのアクセス範囲が定まらない．または関係者間の関係（指導教員の関係など）の使用シナリオを挙げないことがわかった．

4.3 その他

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．

本稿では、LUBM のデータ生成モデルを参考することで、現実情報に近い、災害対策の実際問題にあるシナリオを導入したデータセットを生成するベンチマークツールを実装することを目的とした．

第5章

SIBM - 避難場所情報ベンチマーク ツール

5.1 SIBMのコンセプト

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. 現実に近いデータの特徴や使用シナリオのこと：災害対策のため、実際の災害にすぐ対応できることはとても大切な目標である。
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よって、様々な場面で活用できることが考えられる。

5.2 データ構成の設計

SIBMが2つの情報：避難場所情報、人間（ユーザ）の情報を対象とする。避難場所とユーザのつながりが図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

ユーザが3つの対象がある：アシスタント、避難者、ボランティア。ユーザ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。

Fig. 5.1: 避難場所データ構成

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。

Fig. 5.2: 避難場所全体の構成

また、家族全員が違う避難場所で避難することがある。SIBMでは、そのような関連を追加することが可能である。図3のように、違う避難場所で避難する人々間の家族関係が生成される。

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。本稿では、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. 5.3: 家族関係

Fig. 5.4: 都道府県単位の避難場所構成

5.3 データクラスの定義

SIBMでは、入力したパラメーターによる、データセットの規模や情報の複雑度を適切に生成される。データが生成される際、データ各種を部分的に生成され、最後に設定したパラメーターを導入して、結果を出力する。

関係者情報クラスの詳細設計が、図5.5のようになる。

共有となるクラス：**Person クラス**と**UserType クラス**がある。

- **Person クラス**が、関係者の一般的な情報を持ち、それぞれに対するフィールドを定義する。
- **UserType クラス**が関係者の種類を分ける情報とその種類に対するアクセ

Fig. 5.5: 関係者クラス図

スレベルを持つ.

関係者クラスを定義する時、そのクラスが **Person クラス** を継承する．そして、そのクラスの種類を定義する **UserType クラス** の確定インスタンスを持つことになる．図 5.5 のように、避難者 (**Evacuee クラス**)、アシスタント (**Assistant クラス**)、ボランティア (**Volunteer クラス**) が定義される．

避難場所情報が図 5.6 のように、5.2 に設計したデータ構成を実装したクラス図である．

Fig. 5.6: 避難場所クラス図

- **GeoPoint クラス**が、避難場所の地上位置を定義するクラス．実際に、避難場所の位置検索などに使用されることがある．
- **HazardClassification クラス**が、特定の避難場所が対応できる災害のタイプを定義するクラス. `true` か `false` 値を持ち、避難できるかできないかの意味を持つ．SIBMでは、6つの災害タイプを対象とする：地震・津波・火山・洪水・その他・未確定．
- **ShelterPoint クラス**が、避難場所の一つを定義するクラスである．
- **Prefecture クラス**が、都道府県の情報とそこにある避難場所の情報を定義するクラスである．

5.4 情報リソース

SIBMでは、現実に近い情報を生成する目的を持つ．その目的を実装するために、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を使用した．SIBMが使用する避難場所情報では、自然災害発生時に住民を避難させ、又は避難住民等の救援を行うための施設で、市町村長が指定し地域防災計画等に掲載されている施設（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ることにならない）．SIBMが生成する人間の個人情報、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目標がある．上記の実装を行うためのリソースは以下のように説明する．

- **国土交通省（MLIT^{*}）**による提供した避難施設情報[†]．SIBMが生成する避難場所情報がMLITによる提供した避難施設情報を使用することになった．その情報の特徴が（表5.1）からわかる．

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

^{*}Ministry of Land, Infrastructure, Transport and Tourism

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

- 人間の情報では、**Random profile web API (RPA)**[‡]を参考にしたツールを作成して使用する．RPAで生成した個人情報 は表5.2のようになる．しかし、RPAでは、使用制限があるため、SIBMがそのモデルを参考にして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
ユーザID	54b0ae57a0dcc
firstName	心
surname	真塩
translitFirstName	Kokoro
translitSurname	Mashio
gender	Female
birthday	1993-09-01
age	21
country	Japan
passportNumber	TX1338642
phone	+8152-324-1820
address	452-1179, Kawazoe, Shiraoui-cho Shiraoui-gun, Hokkaido
email	Kokoro.Mashio@e4xample.com
occupation	Animal Control Worker

TABLE 5.2: 個人情報の例

[‡]<http://randomprofile.com/api-for-developers>

第6章

実験・考察

6.1 実装例

5.4で説明したリソースを使用することと、それらから習得したデータをSIBMのデータ交換環境でRDF形式のデータを得る．避難場所情報とそれに対応するRDF述語の詳細が表6.1のようになる．避難場所情報とそこに作業する人の個人情報が図6.2のように出力される．

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 6.1: 避難場所情報の RDF 述語の詳細

Fig. 6.1: 生成時間

避難場所数	生成・挿入時間 (ms)	関係者数	トリプル数	サイズ (MB)
1	4201	399	4411	192.166
10	6654	3417	37753	193.353
20	9905	9679	106795	195.735
50	18560	14239	157435	197.430
100	50471	48871	539187	265.488
200	120410	78264	864110	315.103
500	501903	181929	2009225	499.542
800	1171083	322574	3561120	742.762

TABLE 6.2: 避難場所数に対する生成時間

6.2 考察

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。

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

```

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/P20_102395>
6    sibm:address "千葉県富里市高野 547"^^ xsd:string ;
7    sibm:administrativeAreaCode "12233"^^ xsd:int ;
8    sibm:facilityScale "114"^^ xsd:int ;
9    sibm:facilityType "避難施設"^^ xsd:string ;
10   sibm:hazardClassification [
11     sibm:earthquakeHazard false ;
12     sibm:notSpecified true ;
13     sibm:other false ;
14     sibm:tsunamiHazard false ;
15     sibm:volcanicHazard false ;
16     sibm:windAndFloodDamage false
17   ] ;
18   sibm:name "高野集会所"^^ xsd:string ;
19   sibm:seatingCapacity "-1"^^ xsd:int ;
20   geo:geopoint [
21     geo:latitude "35.7092377518"^^ xsd:double ;
22     geo:longitude "140.346932058"^^ xsd:double
23   ] .
24  <http://lab.ene.im/SIBM/person/54b09c4cb8885>
25    a <http://lab.ene.im/SIBM/thing/usertype/volunteer> ;
26    sibm:profile [
27      sibm:age "36"^^ xsd:string ;
28      sibm:birthday "1978-07-07"^^ xsd:string ;
29      sibm:email "Yoko . Banba@example . com"^^ xsd:string ;
30      sibm:firstName "陽子"^^ xsd:string ;
31      sibm:gender "Female"^^ xsd:string ;
32      sibm:phone "+8147-580-6051"^^ xsd:string ;
33      sibm:surName "番場"^^ xsd:string ;
34      sibm:userID "54 b09c4cb8885"^^ xsd:string
35    ] ;
36    sibm:stayAt <http://lab.ene.im/SIBM/thing/shelterpoint/P20_102395> .
37  <http://lab.ene.im/SIBM/thing/usertype/volunteer>
38    sibm:accessLevel "2"^^ xsd:int ;
39    sibm:type "volunteer"^^ xsd:string .

```

Fig. 6.2: 実装例

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かかる．

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6.1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

実験環境	
Operation System	Windows 8.1 Pro 64 bits
CPU	Intel core i7 3770, 4 cores 8 thread @ 3.4Ghz (Boost- able up to 3.9Ghz)
RAM	16GB
Storage	HDD
Java	1.7.0_71 (64 bits)
Apache Jena	2.12.1
SIBM	v1.0 build 20150113

TABLE 6.3: 実験環境

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、ある範囲以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 6.1 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

結論はかくかくしかじか．結論はかくかくしかじか．結論はかくかくしかじか．結論はかくかくしかじか．結論はかくかくしかじか．結論はかくかくしかじか．

残された課題はかくかくしかじか．残された課題はかくかくしかじか．残された課題はかくかくしかじか．残された課題はかくかくしかじか．

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

Listing A.1: クエリ 1

```
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 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
5
6 SELECT ( count(*) AS ? count)
7 WHERE {
8     ? x sibm:profile ? y .
9     ? y sibm:age ? z ;
10
11     FILTER(? z > " 60 ""^ xsd:integer)
12 }
```

A.1 がお年寄りの問い合わせで検索したクエリである .