

概念モデルによる異なるドメイン間のリソース連携方法の提案と評価

M2015SE013 棚瀬 泰宏

指導教員 青山 幹雄

1. はじめに

自動車のドライバを対象としたコンテキストウェアシステムでは、ドライバの健康状態を測定したセンサデータと電子カルテシステムなどの医療データを連携する必要がある。しかし、これらのデータを含むドメインは、データ固有の構造や表現を持つため、ドメイン間の連携が困難である。

2. 研究の課題

本研究では、ドライバを対象とするセンサデータと医療データを RDF リソースとして表現し、リソース同士を関係付けることによって各ドメインを連携することを研究課題とする。

3. 関連研究

3.1. センサデータの RDF リソース連携方法

住居内に設置されたセンサデータを OWL と RDF スキーマによって、RDF リソースとして表現する方法が提案されているが、異なる対象のドメイン間とのリソース間を連携する方法は提案されていない[2]。

3.2. 共通述語を持たないドメインの RDF リソース連携方法

同一性であるが表記が異なるリソースと、共通述語を持たないドメインの RDF リソースに対して、リソースの属性情報と外部の Linked Data を用いた連携方法が提案されている。しかし、連携方法は外部の Linked Data の情報量とリソースの"同一性リンク"による関係付けに依存している[3]。

4. アプローチ

センサと医療ドメインの RDF リソースは共通述語を持たず、リソースの"同一性"による関係付けが不可能である。また、ドメイン固有のリソース表記や構造に従って記述されているため、各ドメイン共通の関係が発見されにくく、関係付けを行うことが困難である。

本研究では、RDF リソースとして記述されたセンサデータと医療データを連携するために、属性による連携と概念モデルによる連携の2段階の連携方法を提案する。

5. 提案方法

5.1. コンテキスト概念モデル

自動車を運転するドライバを対象としたコンテキストウェアシステムの入力となるコンテキストの集合体として、コンテキスト概念モデルを構築した(図 3)。これは入力となるコンテキストの概念を、一般化した語彙で関係付けしたモデルである。身体情報は、MEDIS(医療情報システム開発センター)が公開している看護実践用語標準マスター[4]の語彙を用いて

構成されている。また、脈拍数と心拍数はセンサから得られる値が同じであるため同義語の関係で記述した。このモデルを用いることにより、コンテキストの項目の中で各ドメインリソースが持つ共通の要素を明確にする。

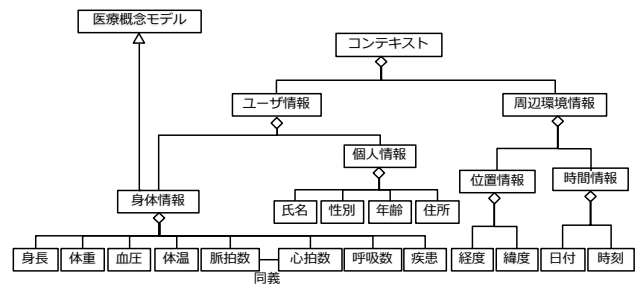


図 1 コンテキスト概念モデル

5.2. 提案プロセス

本研究では、属性による連携と概念モデルによる連携の2段階のリソース連携方法を提案する。

属性による連携では、センサと医療リソースの主要な属性に着目した関係付けを行うことにより連携する。これにより、人名などをキーとした関係付けによってリソース連携可能となる。

概念モデルによる連携では、各ドメインリソースが持つコンテキストの項目を抽象化して統合概念モデルを構築する。構築された統合概念モデルから、各ドメインのモデルの要素間で共通する関係を参照して関係付けを行う。これにより、各ドメインリソースが持つコンテキストの項目に着目した連携が可能になる。図 2 に、本研究の提案プロセスを示す。

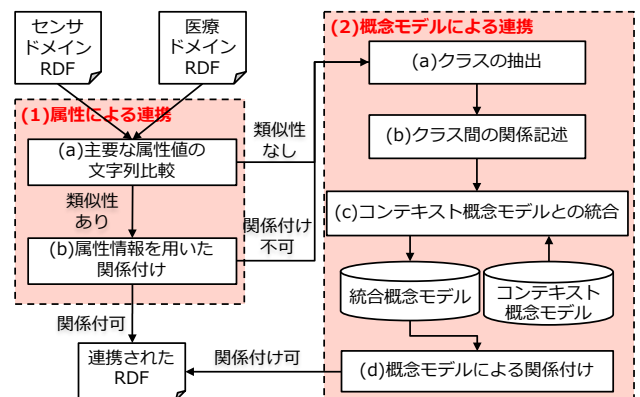


図 2 提案プロセス

(1a) 主要な属性値の文字列比較(属性による連携)

センサデータと医療データの RDF リソースに対して構造を抽出し、抽出した構造を基に各ドメインリソースの主要な属性値の文字列を比較する。主要な属性とは、人名、マイナンバ

一など、ドメイン固有の識別子に依存しない属性のことをいう。文字列を比較し、文字列の類似性がある場合は、関係付けの候補となるリソースを取得する。本研究での類似性があるとは、2つの文字列に対して正規表現を用いて合致可能な範囲と定義する。

(1b) 属性情報を用いた関係付け(属性による連携)

関係付けの候補となるリソースが取得されたら、そのリソースの属性情報を比較する(図3)。比較した結果より、関係付け可能であるならば関係付けの候補であるリソース同士の関係付けを行う。関係付けが不可能な場合は、概念モデルによる連携を行う。

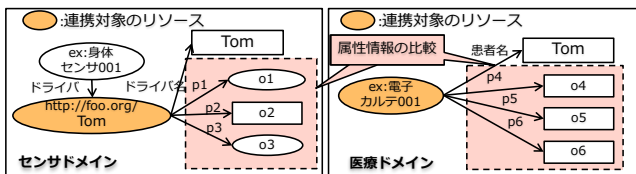


図3 属性情報の比較

(2a) クラスの抽出(概念モデルによる連携)

センサデータと医療データの RDF リソースに対して、rdf:type 述語で関係付けられている基本クラスを抽出する。

そして、リソースの属性情報のクラス(コンテキストの項目)を抽出する。センサドメインでは、コンテキストの項目を記述する述語で関係付けられている目的語を属性情報のクラスとして抽出する。医療ドメインでは、バイタルサインや疾患などのコンテキストの項目を記述する述語で関係付けられている目的語を属性情報のクラスとして抽出する。

(2b) クラス間の関係記述(概念モデルによる連携)

抽出されたクラス間の関係を記述することで、概念モデルへのマッピングを行う。RDF から概念モデルへのマッピング方法を定義した(図4)。

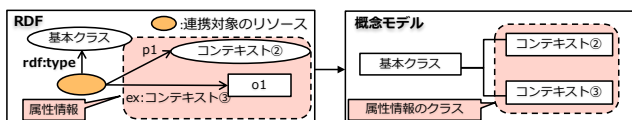


図4 概念モデルへのマッピング方法の定義

(2c) コンテキスト概念モデルとの統合(概念モデルによる連携)

(2b)で関係を記述した各ドメインの概念モデルとコンテキスト概念モデルとの関係を記述し、統合概念モデルを構築する。構築された統合概念モデルを参照することにより、各ドメインリソースのモデルの要素間で共通する関係を特定する。

(2d) 概念モデルによる関係付け(概念モデルによる連携)

構築された統合概念モデルを参照し、各ドメインリソースのモデルの要素間で共通する関係を基に関係付けを行う。

6. プロトタイプの実装

提案方法を支援するシステムのプロトタイプを実装した。プロトタイプの構成を図5に示す。Apache JenaとJavaを用いて SPARQL を利用し、RDF ストアでは Apache Jena Fuseki を利用した。

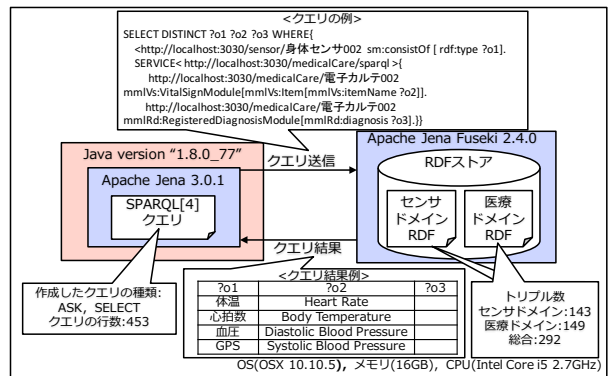


図5 プロトタイプの構成

7. 例題への適用

7.1. 適用する例題

(1) センサドメイン

自動車を運転するドライバを対象に、身体センサで測定可能な心拍数、血圧、体温などの属性、自動車から取得可能な位置情報や測定日、測定時刻などの属性を記述した。

(2) 医療ドメイン

ドライバの電子カルテデータを、MML[1]のサンプルインスタンスを基に記述した。電子カルテデータのリソースでは、患者の血圧や心拍数などのバイタルサインと、患者の疾患名などを記述した。

7.2. 適用プロセス

(1) 主要な属性の文字列が一致する場合(属性による連携)

SPARQL クエリによってリソースの主要な要素である人名を抽出し、「田中太郎」の文字列を持つリソースを発見した。関係付けの候補となるリソース、センサドメインの<田中太郎>と、医療ドメインの<電子カルテ 001>の属性情報を比較した。比較方法として、属性の一致率による同一人物の判定を行った。また、各ドメインにおいて電話番号や誕生日などの属性の記述方法が異なるため、正規表現を用いて部分的なストリングマッチングを行った。その結果、email、電話番号などのドメインによって記述方法が異なる属性の一致、誕生日の属性値の一致から、同一の人物であることを確認した(表2)。

表2 属性情報の比較の一部

属性	センサドメイン	医療ドメイン
氏名	田中太郎	田中太郎
誕生日	08-12	1992-08-12
性別	male	male
国籍	データなし	JPN
住所	データなし	愛知県名古屋千種区 11-1
電話番号	Tel:+09-0-1111-1111	090-1111-1111
e-mail	yamadataro@nise.org	yamadataro@nise.org

同一人物の判定を行った結果、人物のキーとなる属性の一致率により同一人物であることを確認し、所持関係を記述可能な述語 dc:hasPart 述語 による関係付けを行った(図6)。

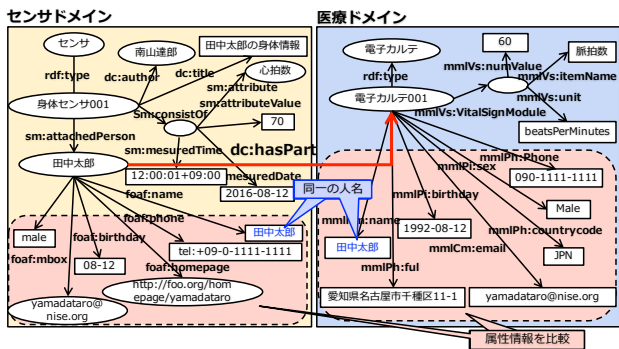


図6 属性による連携の関係付け結果

(2) リソースの言語表記が異なる場合(概念モデルによる連携)

SPARQL によってリソースの構造を抽出し、抽出した結果主要な属性を人名に決定した。人名を特定する述語を用いた SPARQL クエリを実行したところ、「南山花子」という同性同名の文字列が得られたが、属性情報を比較したところ同一人物と判定するための必要な情報が得られなかった(表 3)。

表 3 属性情報を比較した結果の一部

属性	センサドメイン	医療ドメイン
氏名	(データなし)	南山花子
誕生日	(データなし)	1993-01-02
性別	female	female
電話番号	Tel:+080-1234-1234	090-9876-9876

属性による連携で関係付けが不可能だった場合は概念モデルによってリソースのコンテキストの項目に着目した連携を行う。そこで、センサデータと電子カルテデータのリソースのコンテキストの項目(クラス)を SPARQL によって抽出し、クラス間の関係を定義して抽象化した。そして、コンテキスト概念モデルとの関係を記述して統合概念モデルを構築した(図 7)。

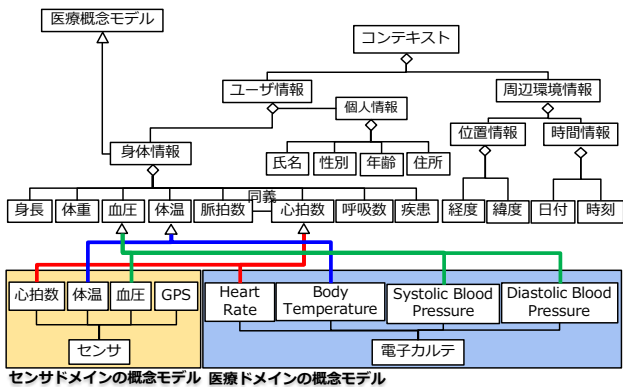


図7 統合概念モデルを構築した結果

構築した概念モデルより、心拍数、体温、血圧のクラスが各ドメインで共通するクラスを特定した。そして、コンテキスト概念モデルの親クラスのリソースと、各ドメインの子クラスの関係から、各ドメインのリソースに対して述語 `rdfs:subClassOf` を用いた関係付けを行った(図 8)。これにより、各ドメインにおいてコンテキストの共通要素である心拍数、体温、血圧のクラスの連携が可能であった。

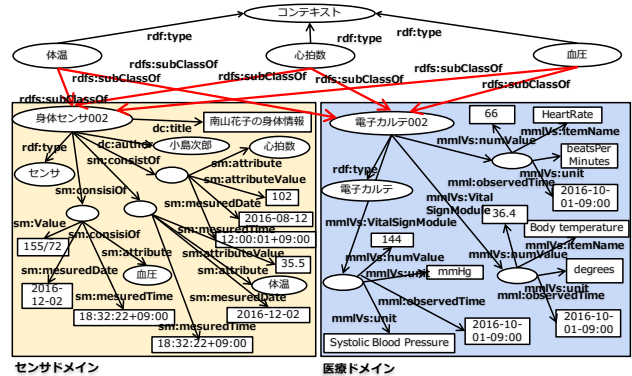


図8 概念モデルによる関係付け結果

(3) 同義語である属性名の名前空間が異なる場合(概念モデルによる連携)

人名を特定する述語を用いて SPARQL クエリを実行したところ、同一の人名が発見できなかった。そこで、各ドメインのリソースのコンテキストの項目を抽象化し、コンテキスト概念モデルとの関係を記述して統合概念モデルを構築した(図 9)。

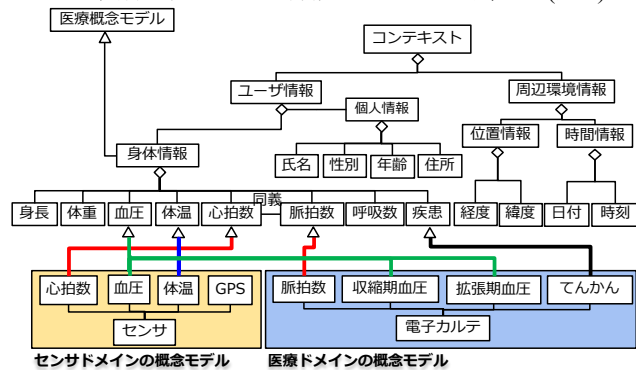


図9 統合概念モデルを構築した結果

図 9 より、各ドメインにおいて、血圧がモデル間の要素の共通する関係として抽出した。また、心拍数と脈拍数などコンテキスト概念モデルで同義語の関係にあるものも共通する要素として抽出した。抽出された関係を基に、概念モデルによる関係付けを行った(図 10)。

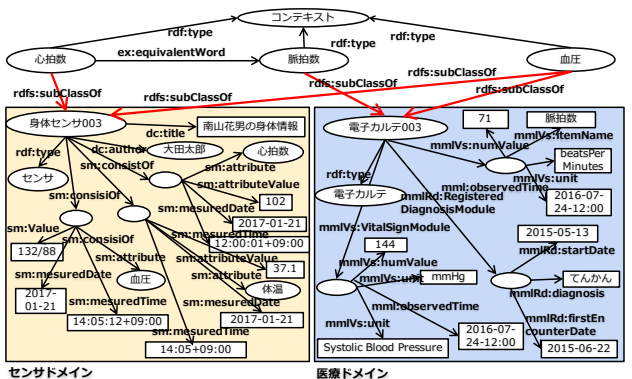


図10 概念モデルによる関係付け結果

7.3. 適用結果

以下、3つの場合についての適用結果を示す。

(1) 主要な属性の文字列が一致する場合(属性による連携)

リソースの主要な属性である人名を SPARQL クエリにより抽出することで、同姓同名の人名を発見し、関係付の候補となるリソースを発見できた。そして、各ドメインの属性情報を比較することにより関係付けを行うことが可能であった。

(2) リソースの言語表記が異なる場合(概念モデルによる連携)

主要な属性の文字列を比較することにより、同姓同名のドライバが対象のリソースを発見できた。しかし、連携対象リソースの属性情報を比較した結果、関係付けに必要な情報が得られなかったため、概念モデルによる連携を行った。センサドメインでは日本語表記、医療ドメインでは英語表記など異なった言語表記で記述されているリソースに対して、概念モデルを構築することにより、心拍数、血圧、体温の3つのコンテキストの共通要素を特定した。抽出された共通の関係を基に関係付けを行うことで、リソース連携が可能であった。

(3) 同義語である属性名の名前空間が異なる場合(概念モデルによる連携)

各ドメインリソースのコンテキストの項目を抽象化して概念モデルを構築し、心拍数と脈拍などの同義語の関係にある要素や、血圧などモデル間の共通する要素の関係を基に関係付けを行い、リソースを連携することが可能であった。

8. 評価と考察

8.1. 例題を用いた提案プロセスの妥当性の評価

(1a) 主要な属性値の文字列比較

主要な属性値を用いた SPARQL クエリにより、同姓同名のドライバが対象のリソースを発見し、連携の候補となるリソースを発見することが可能になった。

(2b) 属性情報を用いた関係付け

属性の一致率によるストリングマッチングにより同一人物を属性値として持つリソースの判別が可能となった。また、dc:hasPart 述語によって、人名リソースと電子カルテリソースの所持関係の関係付けが可能になった。

(2a) 統合概念モデルの構築

属性による連携が不可能だったリソースに対して、基本クラスの抽出と属性情報のクラスの抽出を行い、関係を記述することにより、センサデータと電子カルテデータリソースを抽象化することが可能となった。そして、コンテキスト概念モデルとの関係を記述することによって、統合概念モデルを構築し、連携対象の2つのドメイン RDF リソースの概念モデル間でコンテキストの共通要素を特定することが可能になった。

(2b) 概念モデルによる関係付け

統合概念モデルを参照し、各ドメイン RDF リソースのモデル間で共通する要素の関係を抽出し、関係付けを行うことにより、各ドメイン固有の表記や構造を持つセンサと医療ドメインのリソースに対して連携を行うことが可能となった。

8.2. 例題を用いた2階層のRDFリソース連携方法の評価

(1) 属性による連携

属性を用いた関係付けを行うことにより、センサと医療ドメインのリソースが連携可能となる。また、人名などの属性を

キーとした関係付けを行うことで、同一人物に関する疾患データと動的に変化するバイタルサインデータを連携した分析が可能となると考える。

(2) 概念モデルによる連携

属性による連携が不可能な場合でも、概念モデルを用いることによって、ドメイン固有の構造を持つリソースに対して連携が可能となる。そのため、異なるドメイン間の連携の範囲が拡張可能となったと考える。

8.3. 関連研究との比較

(1) センサ以外の異なるドメイン間のリソース連携

住居内のセンサデータと異なる対象ドメインの RDF との関係性を記述する方法が定義されていなかった従来の方法と比べて、医療ドメインとの連携方法を明確にした点で有意であると考えられる。

(2) 共通述語を持たないリソース連携

異なるドメインリソースに対して、owl:sameAs 述語を用いた関係付けを行っていた従来の方法と比べて、リソース同士の"同一性"の関係付け以外の方法でドメイン間のリソース連携が可能となった点で有意であると考えられる。また、外部の Linked Data の情報量に依存していた従来の方法と比べて、それに依存せず、リソース同士を関係付け可能となった点で、より高度な連携が可能になったため有用であると考えられる。

9. 今後の課題

関係付けられたリソースが、実際に正しい関係付けを行われたリソースであるかの判断は人間が行わなければならない。そのため、その関係付けが正しいかの検証を定量的に行う必要がある。また、プロトタイプは小規模なリソースに対して適用した。そのため、より大規模なリソースにスケールする必要がある。

10. まとめ

センサデータと医療データの RDF リソースは、各ドメイン固有の表現や構造が存在するため連携が困難である。

本研究では、異なる表記や構造を持つセンサと医療のドメイン RDF リソースを連携するために、属性による連携と概念モデルによる連携の2階層の連携方法を提案した。提案方法を例題に適用することにより本研究の妥当性を示した。

参考文献

- [1] MedXML コンソーシアム, Medical Markup Language (MML) Version 4.1.2 規格書, Dec. 2016. <http://medxml.net/MML412j/mml4.html>.
- [2] 野口 博史 ほか, 住居内異種センサの統一的処理のための RDF センサ記述, 電子情報通信学会, ネットワークロボット研究会 NR-TG-2, 2006, pp. 1-6.
- [3] 内海 太祐, 複数コンテキストドメインにまたがる Linked Data を用いたコンテキストモデル構築方法の提案と評価, 情報処理学会第78回全国大会, No. 3J-03, Mar. 2016, pp. 357-358.
- [4] 医療情報システム開発センター, 看護実践用語標準マスター《看護観察編》, Dec. 2016. <http://www2.medis.or.jp/master/kango/kansatsu/index.html>.