

医療情報コードデータベースを用いたメタデータ生成システム Metadata generation system: Using Medical Information Code Database

林 正治† 堀井 洋† 権 仁洙† 吉田 武稔†
Masaharu HAYASHI Hiroshi HORII Insoo KWEON Taketoshi YOSHIDA

1. はじめに

健診データ、検査データ、医用画像データなどの医療データの多くは医療機関ごとに個別の情報処理を行うことが多い。このような状況の中、医療データと関係する医学知識の蓄積と共有に関する研究がある。このような研究課題のひとつとして、それらの知識を広く利用するために、単なるパターンマッチングだけでなく、医療用語シソーラスを用いた類似検索や意味的情報を基にした検索が求められている[1]。この検索には、既存の医療データを意味情報に基づく検索に適した形式で蓄積することが要件となる。

本報告では、意味情報に基づいた検索を可能とするために、医療情報コードデータベース(以下、医療情報DBと呼ぶ)[2][3]からRDF(Resource Description Framework)[4][5]形式のメタデータを生成し、それを用いた医療データの蓄積手法を提案し、その実装について示す。

2. 関連研究

セマンティックWebの適用分野として、医薬品の発見・開発支援[6]、医療データの交換および検索[7][8]などがある。これらの研究において、メタデータに注目すると、文献[7][8]では、新たにRDFを設計し、国際疾病分類(International Classification of Diseases)[9]等の標準化された医療情報コードと関係を持たせることで意味的情報を付与している。

また、セマンティックWebからリレーショナルデータベースに蓄積されたデータを取り扱う研究として、文献[10][11]がある。文献[10]では、リレーショナルデータベースの構造とRDF形式のマッピング言語を提供している。文献[11]では、RDFファイルを検索する独自の言語をSQL言語に変換機構を提供している。しかし、文献[10]の手法では、マッピング言語を予め定義する必要があり、手間がかかる。また文献[11]の手法では、独自の検索言語を用いるため一般的に利用するには不向きである。

本論文では、既存のリレーショナルデータベースからメタデータを生成する手法を提案する。

3. メタデータ生成手法

医療情報DBからメタデータを生成する手順を以下に示す。

- (1) 医療情報DBのスキーマ情報を取得し、RDFSファイルを生成する。
- (2) 医療情報DBからRDFファイルを生成する。
- (3) RDFファイルをRDFデータベースに格納する。

医療情報コードスキーマの作成指針は、医療情報DBのテーブル名を医療コードスキーマのクラス名とし、医療情報DBのデータ項目名を医療コードスキーマのプロパティ名とする。属性については、今回は省略した。また、メタデータを作成する時のリソースURIには、医療情報DBのURIとテーブル名、そして主キー項目を組み合わせて記述する。主キー項目は一意な値であ

るため利用した。また、主キーが存在しないテーブルの場合は列番号を用いた。

- (1) 主キーがある場合のリソースURI

```
jdbc:postgresql://localhost/normaldb?symptom#S000548
```

- (2) 主キーがない場合のリソースURI

```
jdbc:postgresql://localhost/normaldb?interaction#RowNum32683
```

その他の値は、医療情報DBの各項目名から作成したプロパティのリテラル値として記述する。

表 1. 症状措置機序マスタテーブル(SYMPTOM)

name	説明	属性
symptom_code	病状措置機序コード	varchar, primary key
symptom	症状・措置	varchar
action_mechanism	作用機序	varchar

```

1: <rdf:RDF
2:   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3:   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
4:   <rdfs:Class rdf:about="jdbc:postgresql://localhost/normaldb?symptom#SYMPTOM">
5:     <rdfs:comment>symptom</rdfs:comment>
6:   </rdfs:Class>
7:   <rdfs:Property rdf:about="jdbc:postgresql://localhost/normaldb?symptom#symptom">
8:     <rdfs:domain rdf:resource="jdbc:postgresql://localhost/normaldb?symptom#SYMPTOM"/>
9:     <rdfs:comment>symptom varchar(1000)</rdfs:comment>
10:  </rdfs:Property>
11:  <rdfs:Property rdf:about="jdbc:postgresql://localhost/normaldb?symptom#symptom_code">
12:    <rdfs:domain rdf:resource="jdbc:postgresql://localhost/normaldb?symptom#SYMPTOM"/>
13:    <rdfs:comment>symptom_code varchar(200)</rdfs:comment>
14:  </rdfs:Property>
15:  <rdfs:Property rdf:about="jdbc:postgresql://localhost/normaldb?symptom#action_mechanism">
16:    <rdfs:domain rdf:resource="jdbc:postgresql://localhost/normaldb?symptom#SYMPTOM"/>
17:    <rdfs:comment>action_mechanism varchar(1000)</rdfs:comment>
18:  </rdfs:Property>
19: </rdf:RDF>

```

図 1. テーブルのRDFSファイル

4. メタデータ生成システム

提案手法を評価するためにプロトタイプシステムをJava言語で構築した。Java実行環境はJ2SE 1.5.0_06, セマンティックWebライブラリにJena 2.4[12], データベースにはPostgreSQL 8.1.2を用いた。

本研究では、医療情報DBとして厚生労働省保険局が公開している医薬品マスタ[2], 日本医師会が公開している併用禁忌

† 北陸先端科学技術大学院大学 知識科学研究科
School of Knowledge Science, Japan Advanced
Institute of Science and Technology

薬データベース[3]を利用する。医薬品マスタは1つの表に 17, 519 件のデータを格納している。併用禁忌薬データベースは相互作用テーブル、症状措置機序マスタの2つの表から構成され、それぞれ 992 件、56,040 件のデータを格納している。表 1 は症状措置機序マスタのテーブル構成である。

次にメタデータ生成システムの具体的な処理を示す。まず、医療情報DBのスキーマ情報からRDFSを生成する(手順(1))。

- (1) 医療情報DBからJDBCのgetMetadata()を用いてデータベースのスキーマ情報を取得する。
- (2) (1)からgetColumnNames()を用いてテーブルのスキーマ情報を取得する。
- (3) (2)でNULLが許可されていない列を検出し、それを主キーと定める。
- (4) テーブル名をrdfs:Class, 列名をrdf:Propertyと定める。列の付随情報はrdfs:Commentとして定める。
- (5) (4)をRDFSとして出力する。

図 1 に実際に併用禁忌薬データベースの症状措置機序マスタテーブルから変換したRDFSファイルを示す。4 行目から 6 行目までがテーブルSYMPTOMから変換したrdfs:Class, 7 行目から 10 行目までが列symptomから変換したrdf:Property, 11 行目から 14 行目までが列symptom_codeから変換したrdf:Property, 15 行目から 18 行目までが列action_mechanismから変換したrdf:Propertyである。

```

1: <rdf:RDF
2: xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3: xmlns:symptom="jdbc:postgresql://localhost/normaldb?symptom#">
4: <rdf:Description rdf:about="jdbc:postgresql://localhost/normaldb?symptom#S000001">
5: <symptom:action_mechanism>ネオスチグミンはコリンエステラーゼを阻害し、脱分極性筋弛緩剤の分解を抑制する。</symptom:action_mechanism>
6: <symptom:symptom>脱分極性筋弛緩剤の作用を増強する。</symptom:symptom>
7: <symptom:symptom_code>S000001</symptom:symptom_code>
8: </rdf:Description>
9: </rdf:RDF>

```

図 2. テーブルから生成したRDFSファイル

次に、医療情報DBからRDFSファイルを生成する(手順(2))。

- (1) 手順(1)で主キーが取得されていれば、主キーを用いてURIを作成する。無ければ列番号からURIを作成する。
- (2) 手順(1)で取得したテーブル名を用いて、テーブルの全項目を選択する。
- (3) 手順(1)で変換したRDFSファイルを用いて各列の値をリテラル値とし、RDFSファイルを出力する。

図 2 は併用禁忌薬データベースの症状措置機序マスタテーブルから主キーS000001 で選択された行を用いて生成したRDFSファイルの例である。4 行目に主キーを用いて生成したURIがみられる。表 2 にメタデータ生成に要した処理時間を示す。各テーブルの処理時間はほぼ 1 分以内であり、実用的な範囲であると考えられる。

5. まとめと今後の課題

本報告では、既存の医療データが意味検索に適した形式で蓄積されていないことを問題とし、医療データからRDFS形式のメタデータを生成する手法を提案、その実装を示した。

文献[10][11]では、リレーショナルデータベースからメタデータを生成するには時間がかかるとしていた。しかし、提案手法を実装した情報システムによってリレーショナルデータベースをメタデータに変換する手法の実用性を示すことが出来た。

提案手法により既存の医療データから意味検索することが可能となった。今後は、医学知識と医療データの関係付け及び共有手法、そして意味検索システムの構築を行う予定である。さらに、医学知識の共有及び意味検索システムによる診断プロトコル支援の実現を目指す。

表 2. メタデータ変換処理時間

テーブル名	レコード数	項目数	時間(秒)
医薬品マスタ	17519	32	68.118
相互作用テーブル	992	3	2.063
症状措置機序マスタ	56040	3	29.222

謝辞

本研究は、文部科学省知的クラスター創生事業の一部として行われた。関係各位に謹んで感謝の意を表す。

参考文献

- [1] 医療分野における情報技術ロードマップ. 平成 17 年 1 月, 独立行政法人情報処理推進機構.
- [2] 診療報酬情報提供サービス: 医薬品マスタ. 厚生労働保険局.
- [3] 併用禁忌データベース. 日本医師会.
- [4] W3C: Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- [5] W3C: RDF Vocabulary Description Language 1.0: RDF Schema. <http://www.w3.org/TR/rdf-schema/>
- [6] Eric Neumann: Finding the critical path: Applying the semantic web to drug discovery and development. Drug Discovery World, pp.25-33, Fall, 2005.
- [7] Carro S.A., Scharcanski J.: A framework for medical visual information exchange on the web. Comput. Biol. Med., Apr;36(4):327-38, 2006.
- [8] Boulos MN, Roudsari AV, Carson ER: Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the qualified Dublin Core metadata set, Med. Sci. Monit. Jul;8(7):MT124-36, 2002.
- [9] WHO: International Classification of Diseases (ICD), <http://www.who.int/classifications/icd/en/>
- [10] Christian Bizer, Andy Seaborne: D2RQ - Treating Non-RDF Databases as Virtual RDF Graphs, Third International Semantic Web Conference, 2004
- [11] Johan Petrini, Tore Rich: Processing Queries over RDF Views of Wrapped Relational Databases, 1st International Workshop on Wrapper Techniques for Legacy Systems, 2004
- [12] Jena Semantic Web Framework, <http://jena.sourceforge.net/>