

1. 医療情報システムにおけるセマンティックウェブ技術の応用 *Symposium*

笹井浩介

コニカミノルタテクノロジーセンター

1. 背景と目的

ネットワークの普及に伴いインターネット上には医療に関する多くの情報が氾濫しており、医療機関において多くの情報が電子化され蓄積されている。しかし現状ではそれらの情報が必ずしも有効利用されているとはいい難い状況である。すなわち患者や家族が医療知識を得るために、難解な専門書を調べたりインターネットを検索して膨大な情報から信頼できる知識を取捨選択して獲得する必要がある。医療従事者においても専門知識は主に専門書や学術論文といった印刷物を入手して時間をかけて学習することによって専門知識を獲得する必要がある。いままでこのような状態が続いている背景には、情報を抽出する際に情報を利用する側の知識や論理を活用する技術が確立できていなかったことが挙げられる。

このような反省に基づき、最近ではセマンティックウェブを構成するResource Description Framework (RDF)や概念体系(オントロジー)などの技術を医療情報に応用しようという取り組みが提唱され^{1,2)}、研究報告もみられるようになってきた^{3~6)}。日本医療情報学会においても2006年には「医療情報とセマンティックス」をテーマとした学術大会が神戸で開催され大盛況であった。われわれは2002年からこれらの技術に着目して利用する側の知識や論理に基づいた医療情報データベースを構築すると同時に、いったん構築したデータベースを利用者の立場で進化させていくことが可能な情報処理技術について多くの大学や研究機関で連携して研究を行ってきた^{7~9)}。本報告では人間の知識獲得を支援することが可能なデータベース技術と、それらを応用したいくつかの研究成果について述べる。

2. セマンティックウェブ技術の応用

セマンティックウェブは、データに意味(セマンティック)を付与することで、コンピュータで処理できるように記述するための技術であり、1998年にイギリス出身の計算機科学者Tim Berners-Leeによって提唱された。その後、データに意味を付与する表現方法の枠組みについてWeb技術の標準化団体World Wide Web Consortium (W3C)において検討され、セマンティックウェブの標準技術としてRDFを最終承認し、W3C勧告として2004年2月に公開した。

RDFは従来のRelational Data Base (RDB)とはデータモデルが異なり、簡単なルールでデータを構成する

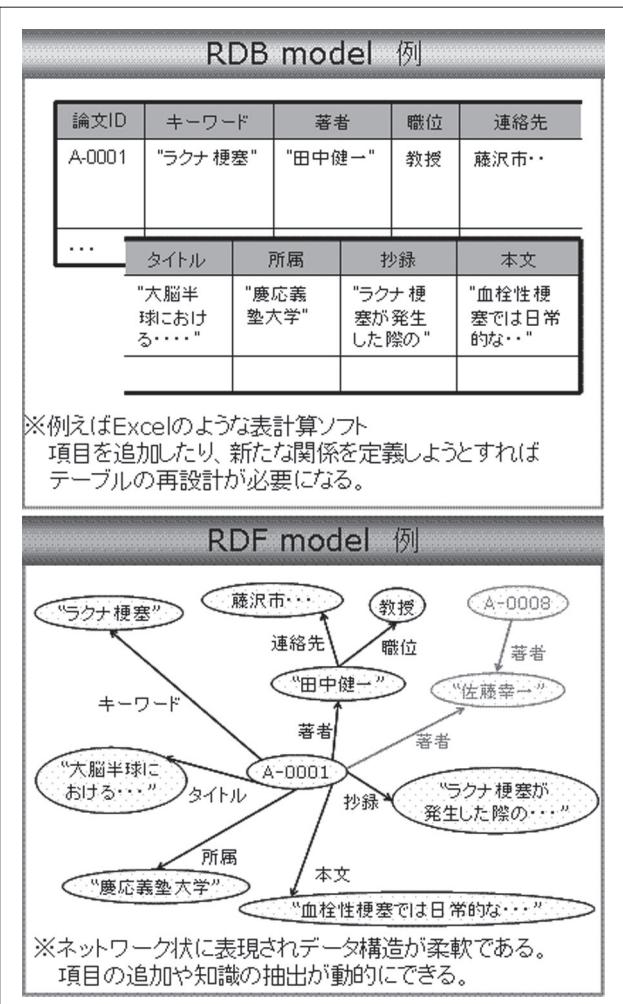


Fig. 1 Comparison between RDB model and RDF model

それぞれのエレメントとその関係を動的に記述していく。したがってデータベースを設計する際に不確実性を有した情報であっても容易にデータベースに追記していくことができる。なお、それぞれのデータエレメントはタグでデータに意味や構造を埋め込んでいくマークアップ言語であるExtensible Markup Language (XML)を用いてデータにセマンティックを付与する。

Fig. 1に論文データにRDFを応用した場合のデータ構造について、従来のRDBとの対比で示す。Fig. 1の上図はRDBの例である。列に論文を形成するデータエレメント、行に論文IDを割付、それぞれのセルにデータを挿入している。論文を追加したりデータエレメントごとに検索してソートすることができる。Fig. 1の下図はRDFの例である。論文IDを中心にデータエレメントとその関係が数珠つながりで蓄積される。も

もちろん同一のキーワードが他の論文に使用されていたり、著者が他の論文を執筆していたり共著であった場合は、その関係も表現される。その結果、全体が膨大なネットワーク状のデータベースとして表現される。データベースから特定のデータエレメントと特定の関係のみを抽出することにより必要な知識が提供される。

3. 応用例と今後の方向性

RDFを応用してネットワーク状に関係を記述したデータベースを構築することにより以下に示すような知識が抽出できる。

論文や学術文献から必要な情報を正しく抽出するためには専門用語を正しく用いることが重要である。しかし専門用語は多くの複合語や同義語が存在すると同時に常に新陳代謝を繰り返している。辞書だけ

では常に新しい言葉やローカルな要素まで含めた特殊な表現を網羅することは不可能である。本システムでは論文を検索する際に利用者から専門用語の知識を吸収することにより、動的に専門用語のオントロジーを生成することができる。さらにこのオントロジーを利用して研究グループやキーマンの抽出が可能である。

一方、所見レポートや退院サマリなどの臨床データを利用して、さまざまな切り口から過去の症例を分析して診断支援を行うことができる。また看護介護における症例や対処方法における経験をデータベース化することにより、ヘルスケア領域に寄与するようなシステムを構築することも可能である。さらに今後は個々のデータベースを連携させることにより、オンラインティーチングや診療シミュレーション、e-learningなどへの応用も期待される。

参考文献

- 1) 高度情報通信ネットワーク社会推進戦略本部(IT戦略本部). 重点計画-2006. 2006: 10.
- 2) 里村洋一. “用語と分類”からオントロジーへ. 医療情報学 2006; 25(6): 377-384.
- 3) 今井 健, 荒牧英治, 梶野正幸, 他. 構文情報と医学用語属性を用いた画像診断所見オントロジーの構築の試み. 医療情報学 2006; 25(6): 395-403.
- 4) 阿部明典, 小作浩美, 相良かおる, 他. 看護リスクマネジメントのための看護オントロジ構築. 医療情報学 2006; 25(6): 431-441.
- 5) 鈴木博道, 清水 昇, 足立和夫. セマンティックWebを用いたMedDRAオントロジーの開発とその利用方法の研究. 医療情報学 2006; 25(6): 447-455.
- 6) 竹村匡正, 黒田嘉宏, 余 直人, 他. 手術手順書からの知識抽出による教育用手術VR環境の要件抽出. 医療情報学 2006; 25(6): 457-462.
- 7) 笹井浩介. 利用者の意図が理解できるデータベース検索システムの開発. PHARM STAGE 2004; 4(6): 33-40.
- 8) 安永 晋, 川上洋一, 笹井浩介. Support Vector Machine (SVM)を用いた自然文読影レポートからの医学的知識の抽出. 医療情報学 2006; 25(6): 405-412.
- 9) 川上洋一, 松村泰志, 笹井浩介, 他. レポートイングシステムにおけるRDFの応用. 医療情報学 2006; 25(6): 421-429.