

セマンティック Web 技術を用いた医療情報検索システム

Ly Bich Lam Ngoc[†] 野崎一徳[†] 伊達 進[†] 下條 真司[‡]

あらまし 近年、医療分野における研究や患者の診断の効率化のために、医療情報の共有・統合の必要性が高まっている。特に、医療画像に対して医師が記述する所見などのデータの共有は診断効率化の上で重要である。しかし、今日、異なる病院間では、医療情報管理データベースシステムが異なり、医療データベース間の情報共有が困難である。例として、データベースのデータ構造が異なるため、同じ対象物を示すデータであっても、検索が困難であることが挙げられる。さらに、医師が患部や所見を記載する際には、同じ意味を示す病名情報であっても、医師によって表現が異なることも多い。そのため、患者履歴や同様の症例をデータベースをキーワードによって検索するのみでは目的とする情報を完全に取得することはできない。本論文では、セマンティック Web 技術を用い、歯科医療のための医療情報検索システムについて述べる。その結果、提案システムにおける病名の表記の差異を解消する検索が実用的な時間で実現できることを示した。

キーワード 医療画像情報, ICD-10, オントロジ

A Searching System for Medical Information by Using Semantic Web Technology

Ly Bich Lam Ngoc[†] Kazunori Nozaki[†] Susumu Date[†] Shinji Shimojo[‡]

Abstract Recently, in medical field, since sharing information for research or diagnosis has become significant, the requirement of medical data integration between various medical units is getting higher and higher. In particular, sharing data such as diagnostic image findings for improving the quality of medical diagnosing is vitally important. However, reaching medical data integration is a great challenge because the medical data management system varies according to the medical units. For instance, the difference of schema - data structure - of databases obstructs the extraction of the same field data from these databases. Furthermore, the using different terms according to medical physicians to describe the same disease in electronic medical chart usually occurs so that using keyword to find patient records or similar cases through various databases hardly gets the full desired information. In this paper, we propose a method of constructing a searching system for dental medical by using semantic web technology. The evaluation in this paper has showed that our proposed searching system is able to solve the variance resulted from using the various terms for the same disease in a practical time.

Keywords Medical image information, ICD-10, ontology

1 はじめに

近年の医療情報の電子化により、さまざまな診療情報が多様なデータベースに蓄積されつつある。これにより、病歴などへの同一患者のデータアクセスが容易になってきた。しかし、所見が類似した症例を容易に検索したいという要求に対し、現在このような要求を満たした病院情報システムは数少ない。

さらに、医療カルテの電子化とともに、病院間での情報共有に対する要求が高まっている。患者が過去に複数の病院で診療を受け、現在の診断に患者の履歴が必要な場合は、現在診療を受けている病院と過去に診療を受けていた病院との間で情報交換が必要となる。また、医師が症例研究や疫学調査を行うためには、複数の病院の情報から類似症例を検索する必要がある。これらの要求

は、医療情報の統合により実現可能と考えられている。しかし、実際には病院によって医療情報管理システムが異なり、医療データベース間での情報共有は困難である。

このような医療データベース間での情報共有において生じる問題として、病院間でのデータ形式の差異、データ構造の差異および表現の差異などがあげられる。システム間でのデータの形式（データフォーマット）が異なれば、その情報を適切な形式に変換する必要がある。同じ形式で記述されたデータでも、データ構造が異なり構造間（スキーマ間）との関連が定義されなければ、異なるデータ構造から同じデータを抽出することは不可能である。また、表現の差異の問題に関しては、同じ意味の病名であっても、医師によって表現が異なることが多く、患者履歴や同様の症例をデータベースからキーワードにより検索しても、目的とする情報を完全に取得することはできない。

本論文では、歯科における診断画像を用いた診断プロセスに着目し、その表現差異の問題を対象とし、歯科医療のための医療情報検索システムについてまとめる。具体的には、セマンティック Web 技術を応用し、国際標準

[†] 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University

[‡] 大阪大学サイバーメディアセンター
Cybermedia Center, Osaka University

疾病分類に基づくオントロジを導入し、歯科医療における病名の表現差異を解決する検索システムを構築する。特に、本システムでは、データ構造の差異も考慮し、近い将来異なるデータ構造を有するデータベース間で、そのデータ構造間を容易にマッピング可能とするために、RDF (Resource Description Framework) 言語によるデータ構造記述を採用する。

以下、2節では病院間でのデータベース統合に関する課題について整理する。3節では本研究で提案する歯科医療のための医療情報検索システムの概要と設計についてまとめる。4節では、具体的なシステムの構築と実装について述べる。5節では、検索シミュレーションで構築したシステムの評価を述べ、6節で本論文をまとめる。

2 医療情報検索システム実現への課題整理

2.1 標準化への対応

病院での情報共有を目指し、取り上げた表現の差異の問題を解決するために、現在類義語辞書から類推を行い、表現を統一して扱える機構が構築されている。病名表記に関しては、WHO (World Health Organization) が ICD-10 (International Classification of Diseases, 10th revision) という統一表記を提案している [4]。患者情報の表記に関しては、医療情報交換のための標準規約である HL7 (Health Level7) や欧州標準化委員会である CEN (Comite Europeen de Normalisation) 等が病院情報交換プロトコルとして存在する。また、画像情報の取り扱い、CT (Computed Tomography) や MRI (Magnetic Resonance Imaging) など撮影した医用画像の形式と、それらの画像を扱う医用画像機器間の通信プロトコルを定義した標準規格の DICOM (Digital Imaging and Communication in Medicine)、または医療画像データを DICOM 規格準拠で統一的に保管管理するマルチベンダ・マルチモダリティ対応のシステムである PACS などの標準が存在する [1]。

これらの標準は主に病院で行われる画像診断レポート等と関連付けられ、医師への情報提供に利用されている画像メタデータの標準である。しかし、すでに存在するデータベースから、それら標準とされる形式への移行や、国や地方によって異なる制度の問題から、より柔軟な情報交換方式が求められている。すなわち、各々のデータベース形式と環境、制度等に合致しつつ、データの構造や表現の差異を解消し、他の医療データベースと情報交換可能なデータベース管理システムが必要不可欠である。

2.2 表現差異の解消

同じ意味の病名であっても、医師によって表現が異なることが多く、患者履歴や同様の症例をデータベースからキーワードにより検索しても、目的とする情報を完全に取得することは容易ではない。例えば、「虫歯」と「う蝕」は同じ意味を表す病名であるが、所見を記述するときに、歯科医師によって「虫歯」と「う蝕」のいずれを

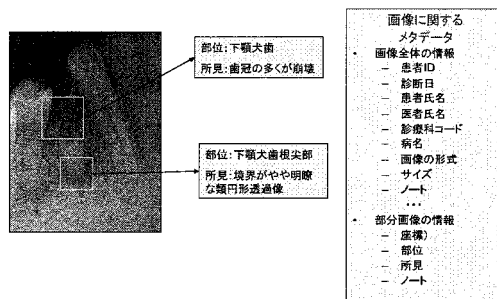


図1 画像に関するメタデータ
(X-ray 像は [2] から引用)

使用するか異なる。そのため、歯科医師が「虫歯」に關する症例をキーワードで検索する場合、「虫歯」の所見のみ見つけるが「う蝕」の所見は見つけることができず、医師の診断プロセスに支障が生じうる。

2.3 データ構造の差異

データの構造であるスキーマが異なるため、同じ対象物を示すデータであっても、抽出が困難となる場合がある。例えば、患者名のデータについて単に名前という一つの項目のデータ構造と名字、名前の二つ項目に分け、患者氏名 (名字, 名前) という構造もある。このような場合、名前の項目を検索対象とすると前者は患者の氏名を抽出できるが語後者は名字はなく、名前のみしか抽出できない。この問題を解決するためにはデータ構造間のマッピングを行うことが必要であり、構造間の関連が定義されれば構造の差異の問題を解決することが可能だと考えられる。

3 提案システム

3.1 システム概要

本論文では、2節で述べた3つの課題のうち、表現差異を解決する歯科医療のための医療情報検索システムについて述べる。また、データ構造の差異の問題について検討をまとめる。

図1は本研究で対象とする歯科医療で取り扱う診療データとそのメタデータについて示したものである。歯科医師は患者を診察する際に、画像を分析し、各患部の周囲にマークを付け、その患部に関して部位名と所見を付記する。この診断手順を考慮し、システムでは医師が患部に記述した各患部情報を画像診断における部分画像情報として管理する。ここで、診断画像においては患者の番号、氏名、医師の氏名、診療科コード、病名、日付、画像形式、サイズ等の情報、画像の患部においては領域、患部の部位と所見の情報が含まれている。

図2は提案システムの概要を示したものである。本提案システムは、ユーザインタフェース部と、病名オントロジデータベース、診断画像メタデータデータベース、診断画像データベース、管理サーバの5つの部分から構

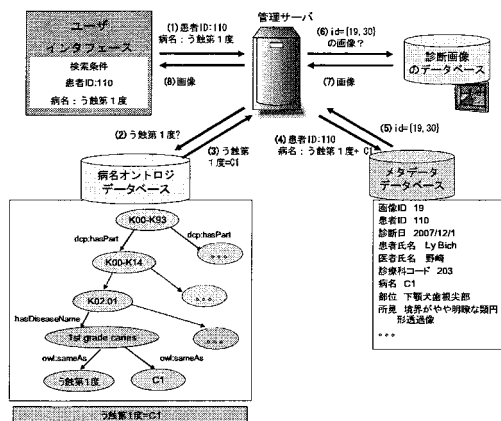


図2 画像検索フロー

成される。管理サーバはシステム全体の動作を指示し、ユーザ入力を受け付けるユーザインタフェースから検索や画像登録等のユーザの要求を受け取り、要求に対応する処理を行う。病名オントロジデータベースは表現差異を解消する役割を担当する。この病名オントロジデータベースは標準病名とその同義語を格納しており、管理サーバからの病名の検索クエリに対して、その病名の同義語を抽出する。一方で、診断画像データベースは診断画像を格納し、管理サーバからの画像 ID 検索により、このデータベースから画像を抽出する。また、診断画像メタデータベースは画像に関するメタデータが格納しており、画像の検索時、このメタデータベース内のメタデータが利用される。管理サーバがメタデータベースに検索を行う際、このメタデータベースは条件を満たした画像メタデータを見つけるとその画像の ID を抽出する。

下記、ID110 番の患者の「う蝕第1度」の診断画像を検索するため、「患者 ID : 110, 病名 : う蝕第1度」という検索条件を入力した場合を例にとり、システム動作について説明する。

- ステップ1 サーバはインタフェースからその要求を受け取った後、検索条件を分析し、病名条件有を確認する。
- ステップ2 「う蝕第1度」の病名の同義語を病名のオントロジに問い合わせをする。
- ステップ3 病名のオントロジから「う蝕第1度 = C1」の病名同義語がサーバに返される。
- ステップ4 オントロジから返した2つの同義語をそれぞれ患者 ID=110 の条件と合わせて、「病名 : う蝕第1度, 患者 ID : 110」と「病名 : C1, 患者 ID : 110」という2つの検索クエリを生成し、それぞれがメタデータのデータベースに問い合わせる。
- ステップ5 2つの検索クエリの条件に対してどちらか満たすメタデータを検索し、結果としてメタデータベースから「19, 30」という2つの画像 ID 番号がサーバに返される。

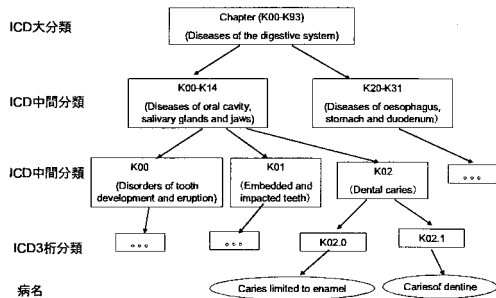


図3 英語版 ICD-10 の例

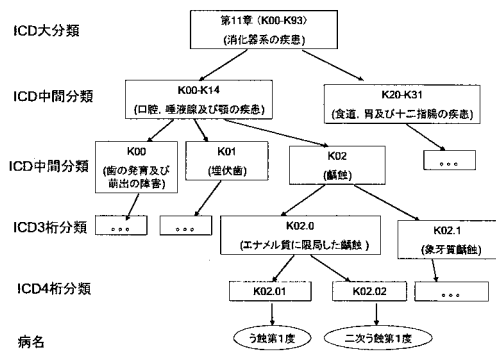


図4 日本語版 ICD-10 の例

- ステップ6 サーバが「19, 30」画像を画像データベースに問い合わせる。
- ステップ7 画像のデータベースから ID と相当する画像がサーバに返される。
- ステップ8 管理サーバが検索の結果画像とメタデータを表示する。

以下では、このような診断画像検索を実現する際、中核的な役割を担う病名オントロジデータベース、診断画像情報データベースについて詳細に述べる。

3.2 病名オントロジ設計

2.2 節で述べた課題である病名の表現差異を解決するためには、標準的な病名、およびその別名の管理が必要不可欠となる。さらに、病名をキーとして病名の表現差異を吸収し、同一の病気に関する診療データへのアクセスを実現するためには、ある病名が入力された際に入力病名と同意の病名を検索可能とするデータベースシステムが必要である。

本研究では、このような視点から、医療分野における標準化された病名と、他言語サポートを考慮し、ICD-10 [6, 3] を採用し、病名オントロジデータベースを実現する。ICD-10 は、先述の通り、WHO による標準化された疾病および関連保健問題の国際統計分類コードであるが、本研究ではその第 10 版である ICD-10 の英語版と日本語版を利用する。

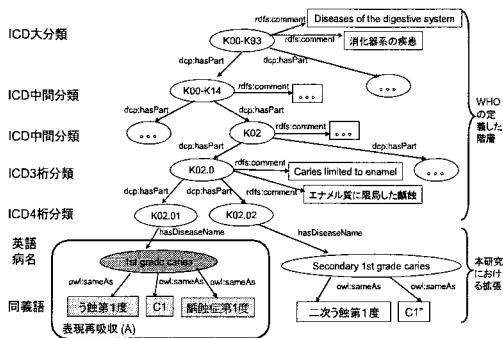


図5 ICD-10の階層を拡張した疾患分類オントロジ



図6 クラスの定義

ICD-10には病名が複数分類コード階層にカテゴリ化され定義される。日本語版 ICD-10 は英語版 ICD-10 と比較し一つ階層が多い。図3と図4はそれぞれ英語版と日本語版 ICD-10 の階層の例であり、大分類 K00-K93 の消化器系の疾患 (Disease of the digestive system) を表す。2つの図から日本語版 ICD-10 は英語版 ICD-10 の3桁分類を拡張して病名をより詳しく分類していることがわかる。

これらの ICD-10 の階層をデータ構造として表現し、さらに標準病名を定義するために本研究は OWL 言語により病名オントロジを構築する。ここで、実際に XML により、データの構造を記述することも可能であるが、XML には親子のデータ構造関係しか定義できないためデータの関係を精密に定義できず、また拡張性に欠けるという問題がある。近い将来、柔軟に拡張可能な機能を追加し、また将来データ構造関係を精密に定義できるように、OWL 言語を選択した。

提案システムにおける病名オントロジでは、図5に示すように日本語版 ICD-10 の5階層をベースに拡張し、日本語病名間を英語の病名で関連付ける。また、英語病名は複数同義語の病名と関連付ける (図5中 (A))。これにより、画像を検索する際に、病名が含まれた場合、病名オントロジを参照し病名の同義語を抽出し、それらの同義語は病名項目条件としてさらにメタデータベースの検索を行うことによって、表現の差異を吸収し、画像を検索可能なシステムを実現可能となる。

以下、OWL 言語 [8] による病名オントロジデータベースのデータ設計の詳細について述べる。図5に示した疾患分類オントロジを表現するために、まず図6に示すように Code クラスを定義する。各階層コードにおいて ICD 大分類、ICD 中間分類 (2 層)、ICD3 桁分類、

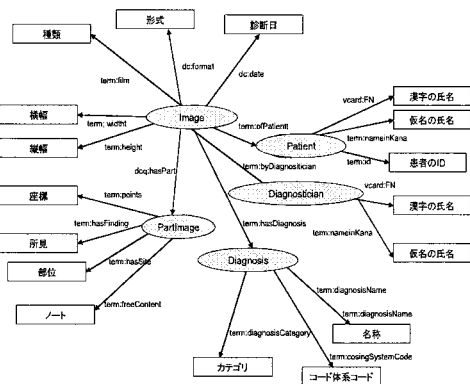


図7 画像メタデータの設計

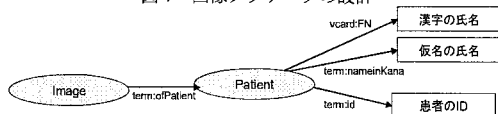


図8 患者データのノード

ICD4 桁分類についてそれぞれ Chapter, BlockCode1, BlockCode2, ICDCode3, ICDCode4 クラス名とし、Code クラスのサブクラスとして定義する。また、標準病名に関してはクラスは Disease クラスを定義し、さらに同義語の病名に関して DiseaseName クラス名にし、Disease クラスのサブクラスとして定義する。

定義したクラスにおいて、それらのデータの関連を Property により表現する。Property について RDF, OWL と Dublin Core [9, 10] がそれぞれ定義した rdfs:comment, owl:sameAs と dcq:hasPart の property を適用し、さらに hasDiseaseName という property を新しく定義する。rdfs:comment は Chapter, BlockCode1, BlockCode2, ICDCode3, ICDCode4 クラスに属するノードに関する内容を説明するために使用する。また、dcq:hasPart は上コードと直下のコード階層を繋げるための Property であり、含む関係を表す。さらに hasDiseaseName は ICD4 桁分類のノードから英語病名ノードを結びつけるための Property である。さらに英語病名ノードを複数の日本語病名の同義語を owl:sameAs に通じて定義する。病名オントロジではコードや病名などはクラスのインスタンスとして定義される。さらにこのオントロジは病名だけではなく分類の内容などを英語と日本語の両方で検索できるように rdfs:comment に通じて日本語と英語の分類内容も導入した。

3.3 画像情報のデータ設計

診断画像の管理において、画像のメタデータとして患者の番号、氏名、医師の名前、診療科コード、病名、日付、画像形式、サイズ等の情報、画像の患部においては領域、患部の部位と所見の情報を管理する。画像のメタデータは図7に示されるように合計17個の項目を含む

表 1 Property の定義

Property	説明	定義域	値域
dc:date	診断日	任意	リテラル
dc:format	画像の形式(jpg, bmp 等)	任意	リテラル
dcq:hasPart	部分画像ノード	任意	任意
term:byDiagnostician	診断医者ノードにつながる	Imageクラス	Diagnosticianクラス
term:codingSystemCode	病名のシステムコード	Diagnosisクラス	リテラル
term:diagnoseName	病名	Diagnosisクラス	リテラル
term:film	医学画像の種類(CT, MRI等)	Imageクラス	リテラル
term:freeContent	ノート	任意	リテラル
term:hasDiagnosis	診断病名ノードに結ぶ	Imageクラス	Diagnosisクラス
term:hasFinding	部分画像の所見	Partクラス	リテラル
term:hasSite	部分画像の部位	Partクラス	リテラル
term:height	画像の高さ	Imageクラス	リテラル
term:id	管理id	任意	リテラル
term:ofPatient	患者ノードに結ぶ	Imageクラス	Patientクラス
term:nameInKana	カナの氏名	任意	リテラル
term:points	部分画像を持つ座標	Partクラス	リテラル
term:width	画像の幅	Imageクラス	リテラル

が、これらのいかなる項目に対しても検索可能とする必要がある。これらのメタデータにデータ構造を有しない場合、検索効率が低く検索に多大な時間を有してしまう。例えば、画像 ID を用いて、患者の名前と患者の ID を検索する時、相当するメタデータには患者の名前と患者の ID を抽出するために、すべての 17 個項目の中で、患者の名前と患者の ID の項目にアクセスする必要がある。反対に、患者の項目を親項目として、名前と ID 子項目を持つデータ構造を持っていると、患者の項目を検索し、その患者の項目から名前と ID 項目を辿るだけで済むため、データ構造を持たない場合に比較して、検索時間を減らすことができる。

このような観点から、本論文では画像メタデータの階層化により、構造を持つデータベースとして効率的に検索可能なデータベースを構築する。また将来の複数病院間でのデータベース統合も考慮し、OWL 言語への親和性が高い RDF 言語を用いて画像メタデータを記述する。

図 7 は設計した診断画像データベースのメタデータの階層を表現したものである。画像データの階層化に関して、画像ノード (Image) を最上位の階層のノードとして、直下の階層に画像メタデータを患者 (Patient)、医者 (Diagnostician)、診断結果 (Diagnosis) そして部分画像 (PartImage) の 4 つ子ノードに繋がる。また、これらの 4 つノードそれぞれが直下の階層に複数の子ノードに繋がる。例えば、図 8 は患者のノード (Patient) の構造を示し、この患者ノード患者の漢字氏名、仮名氏名と患者 ID という 3 つの子ノードを持つ。

この設計にあたって、RDF スキーマのクラスと Property の定義が必要である。以下、本システムで設計した診断画像データベースのメタデータについて詳細を説明する。まず RDF スキーマにおいて、上記説明した階層を表現するために、画像 (Image)、患者 (Patient)、医者 (Diagnostician)、診断された病名 (Diagnosis)、部分画像 (PartImage) というクラスを定義する。

クラスを定義した上で、Property を定義する。表 1 は左列が Property を示し、各行でそれぞれの Property についての説明、定義域、値域を示している。各 Property の

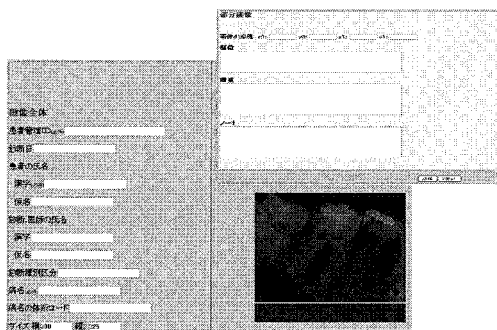


図 9 画像登録

定義域と値域はクラスまたはデータ型で定義される。本システムで定義した Property は「term:」という接頭辞を付けて表現する。そして、メタデータを記述する標準仕様である Dublin Core[9] と vCard[11] によって、定義された語句 dc:date, dc:format, dcq:hasPart, vcard:FN を使用する。ここでは“dc:”、“dcq:”と“vcard:”それぞれは Dublin Core, DublinCore 精密化要素 [10] と vCard の接頭辞である。次に、term:diagnoseName と term:ofPatient を代表として説明する。term:diagnoseName は病名を示す Property である。この Property は RDF トリプル (Subject, Property, Object) で表現し、例えば (A, term:diagnoseName, “虫歯”) なら、A の病名は虫歯であるという意味になる。また、同様に、term:ofPatient の定義域は Image クラスと値域は Patient クラスであるため、トリプルは (X, term:ofPatient, Y) で表現すれば、X, Y はそれぞれ Image クラスと Patient クラスに属することになる。

4 システムの構築と実装

■開発環境 本研究では提案システムを実装するために Java 言語を使用した。前節で述べたように画像に関する情報は RDF 文書、また病名のオントロジは OWL 文書で管理する。RDF, OWL 文書を操作するプログラミング言語としては Java, Perl, C, Lisp などの複数言語が利用できる [7]。その中でも Java 言語で RDF 文書を操作するためのツールを多く揃えている [14]。JenaAPI には ARP RDF/XML パーサ, RDF/XML ライタ, RDQL+SPARQL Query 言語 [13], DAML サポート, persistent storage などの機能が含まれる。本研究ではこのようなツールの整備状況を考慮した上で Java 言語を本システムの開発言語として選択した。

データベースソフトウェアとしては JenaAPI がサポートする MySQL を採用した。また、本研究では本システムを Web アプリケーションとして実装し、サーバサイドプログラムの開発環境としては RDF の操作と同じく Java を採用し、JSP とサーブレット技術を用いて開発した。

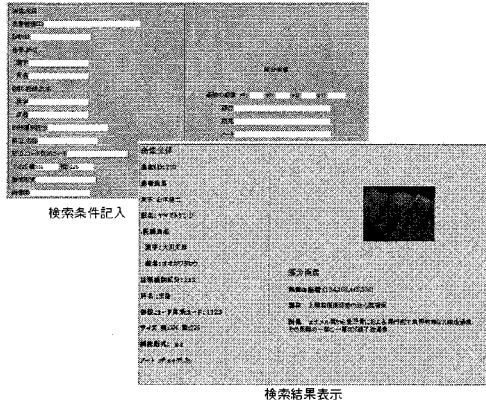


図 10 画像検索

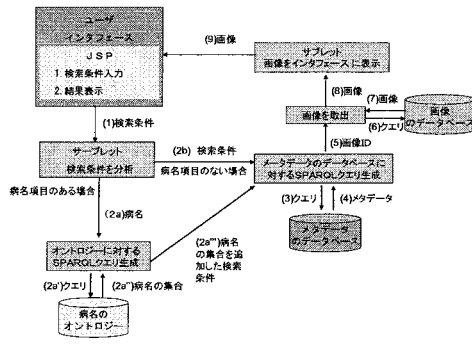


図 11 検索フロー

■医療画像管理システムの構築 本システムは図 9 と図 10 に示す Web インタフェースを提供する。図 9 は画像管理システムにおける画像登録ページである。このページにおいて、ユーザは画像に関する情報と診断した部位に関する情報を付与し、画像の登録処理を行う。また、画像検索を行う際には、図 10 に示すページのように、検索条件を入力し、検索を行う。

本システムにおける、これら画像登録と検索操作実施時の実装の詳細を以下に記す。

登録作業 画像と医師によって画像に付与された情報は JSP の登録ページからサーブレットに渡される。データを受け取ったサーブレットは、画像に対し、システムで使用する画像の ID を生成する。次に画像に付与された情報を分析し、JenaAPI によって、3.3 節で設計した画像情報モデルに従ってデータモデルを作成し、メタデータデータベースに保存する。ここで、メタデータのファイルは画像と同じ ID により管理され、画像ファイルは「imgID」という名称で画像データベースに格納される。

検索作業 本システムでは 2 つの SPARQL クエリを生成する機能を有しており、一つはオントロジに対する SPARQL クエリを生成する機能であり、もう一方はメタデータのデータベースに対する SPARQL クエリを生成

する機能である。これらの 2 つの機能を、それぞれオントロジクエリ関数とメタデータクエリ関数と呼ぶことにする。これらの検索クエリがどのように利用されるかを示したものが図 11 の検索フローである。まず、検索条件が JSP の検索ページからサーブレットに渡される (図 11 中 (1))。サーブレットはその条件項目を分析する。画像病名項目がなければ、その検索条件を引数としてそのままメタデータクエリに渡す (図 11 中 (2b))。一方、病名項目があれば、オントロジクエリ関数でその病名のクエリを生成し、病名のオントロジの参照を行う (図 11 中 (2a) (2a'))。この関数はクエリに使用した病名自身とその病名の同義語を含んだ病名の集合を結果として返す (図 11 中 (2a''))。次にその病名の集合を病名項目の条件とし、検索ページから入力された他の項目の条件と合わせ、メタデータクエリ関数に渡す (図 11 中 (2a'''))。メタデータクエリ関数は与えられた引数を用いてメタデータのデータベースを参照し、検索にマッチする画像 ID を結果として返す (図 11 中 (3) (4) (5))。そして、返された画像 ID を引数として画像を取り出す関数に渡し、実際の画像を得る (図 11 中 (6) (7) (8))。このメタデータクエリに関するステップでは、検索結果にマッチするすべてのデータモデルはメタデータのデータベースから読み出し、総合モデル (Union Model) を作成する。そして、この総合モデルに SPARQL クエリを掛けることによって画像 ID を得る。実際の画像が画像データベースから取り出されると、サーブレットはそれを Web インタフェースに表示する (図 11 中 (9))。

■オントロジ構築 オントロジの構築には代表的なオントロジのエディタである Protege[12] を使用した。オントロジは前節で設計した病名のオントロジモデルに沿って構築した。まず、各クラスと Property を定義し、英語病名と日本語病名を導入した。本実装は歯科の病名を対象として、ICD-10 の K00-K14「口腔、唾液腺及び顎の疾患」病名を使用した。それぞれの病名の同義語は MEDIS-DC[3] を参考に、病名の同義語を定義した。

5 システムの評価

本研究において構築した歯科医療のための医療情報検索システムについて検索時間に関して評価を行った。ここでは、画像のメタデータをデータベースに蓄積し、データ量を変化させ、その検索にかかる応答時間の測定および検索正確性の確認を行った。

■評価環境 本システムの評価を行うために、Pentium 4 2.0GHz, 512MB Memory, WindowsXP が搭載された IBM Laptop PC を使用した。

5.1 検索時間の検証

検索時間を評価するために、病名オントロジの検索時間と画像メタデータの検索に関して、それぞれ次のように応答時間の測定を行った。

1. 病名オントロジの検索時間

オントロジの検索時間の検証においては、566 個の

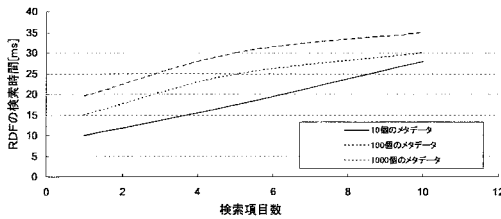


図 12 メタデータ検索時間

病名を保持している病名オントロジの検索時間を測定し、このとき、病名オントロジを用いた 100 回の検索平均時間は 23[ms] であった。システム内に保持する病名数が増えるにつれて、オントロジの検索に要する時間も増加すると考えられるが、病名数が今後爆発的に増える可能性は考えられないことから、この点における本システムの性能は十分実用的だと考えられる。

2. メタデータの検索時間

メタデータの検索時間の検証においては、メタデータデータベースに登録されるデータの個数と、検索に使用する項目数を変更させ、それぞれの条件におけるシステムの応答時間の変化を測定した。図 12 は横軸に検索に使用した項目数を、縦軸に RDF の検索時間をとったグラフである。本評価ではデータベースに登録されているメタデータが 10, 100 と 1000 個の場合、それぞれにおいて評価した。グラフではメタデータの個数の違いは線種で示している。いずれのメタデータ数についても、検索項目数増加にともない、検索に要する時間が増加していることがわかった。また、メタデータ数増加に対し、検索時間にはあまり変化がなく、10, 100, そして 1000 個のメタデータの検索に要する時間のそれぞれの差は 5[ms] 以下であった。一方で、1000 個メタデータで 10 項目のクエリの検索時間は 35[ms] 以下であった。これらの結果から、1000 個程度のメタデータの検索には実用上耐えうる性能が得られると考えられる。

5.2 検索結果の正確性の検証

前節で述べた検索手法を用いたシステムの正確性を、Web インタフェースの動作を含めて検証するために、Web インタフェースから歯科のカルテに関する 20 個のデータの登録と検索を行った。ここで登録した 20 個のデータは予め、人間の手によって評価済みであり、ある検索条件に関してどのような検索結果が出されるべきか把握済みである。システムの正確性の検証では、Web インタフェースから入力した検索条件に対して、予め想定された検索結果が正しく導かれるかに関して、歯科医師の立会いのもと検証した。その結果、病名によるオントロジを用いた検索が正しく実行できていることが確認できた。

6 まとめ

本論文では、セマンティック Web 技術を用い、歯科医療のための医療情報検索システムについてまとめた。提案システムでは、国際標準疾病分類 ICD-10 に基づき OWL で記述された病名オントロジデータベースと、RDF で記述された診断画像のメタデータデータベースを中核とし、病名の表記の差異を解消する検索を可能にする。また、本論文の評価では、提案システムにおける病名の表記の差異を解消する検索が実用的な時間で実現できることを示した。今後は、残された課題の一つである、複数病院間に存在するデータ構造差異の問題に取り組んでいきたいと考えている。

謝辞

本システムの構築において、ご助言頂いた大阪大学歯学部付属病院の玉川裕夫准教授に感謝する。

文献

- [1] 日本医療情報学会: “医療情報,” 医療情報システム編, 篠原出版新社, pp.203-230 (May 2006).
- [2] 山本浩嗣, 小林馨: “歯科放射線の臨床診断: 画像診断と病理概説,” 永末書店, pp.14 (February 1997)
- [3] 財団法人医療情報システム開発センター (MEDIS-DC): “ICD-10 対応電子カルテ用標準病名マスター,” 財団法人医療情報システム開発センター (2002).
- [4] World Health Organization: “International statistical classification of disease and related health problems, Tenth Revision (ICD-10),” (1992).
- [5] World Health Organization (WHO): “International Classification of Diseases (ICD),” <http://www.who.int/classifications/icd/en/>.
- [6] “ICD-10 国際疾病分類第 10 版,” <http://www.dis.h.u-tokyo.ac.jp/byomei/ICD10/>.
- [7] Resource Description Framework (RDF), <http://www.w3.org/RDF/>.
- [8] Web Ontology Language (OWL), <http://www.w3.org/TR/owl-features>.
- [9] Dublin Core, <http://purl.org/dc/elements/1.1/>.
- [10] Dublin Core 精密化要素, <http://purl.org/dc/terms/1.1/>.
- [11] Representing vCard Objects in RDF/XML, <http://www.w3.org/TR/vcard-rdf>.
- [12] The Protege Ontology Editor and Knowledge Acquisition System, <http://protege.stanford.edu/>.
- [13] SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>.
- [14] Jena A Semantic Web Framework for Java, <http://jena.sourceforge.net/>.