

薬剤師向け学習支援システムの開発(第1報)

鍋田 啓太[†] 木村 昌臣[†] 大倉 典子[†] 土屋 文人[‡]

[†]芝浦工業大学工学部 [‡]東京医科歯科大学歯学部附属病院

1 研究背景と目的

医薬品に関わる医療事故は、様々な要因によって引き起こされている。実際に報道されたものだけでなく、ヒヤリハット事例として報告されているものも含めると膨大な数となる。そういった事故を未然に防ぐ方法の1つとして、学習支援システムを用いた薬剤師の効果的な学習というものが求められる。そのためには、薬剤師が学習すべき医薬品情報を収集する必要がある。しかし、既存の医薬品データベースでは、主に検索を目的として作られたものであるため、新たに学習支援システムに利用可能なデータベースを構築しなければならない。したがって、医薬品の正確な情報が記載されている添付文書から情報を取り出し、学習支援システム等のシステムの利用に適したデータベースを構築する仕組みが必要となる。そこで、添付文書に含まれるデータ構造の解析を行い、それに基づいたデータ抽出プログラムの開発を行った。

2 研究内容

2.1 添付文書が持つデータ構造の問題点

添付文書には各医薬品の基本的な情報が記述されている。しかし、この記載内容は情報処理向きに作られていない。

第一に、項目の統一性がないという点が挙げられる。基本的な項目は記載要領に従っているので統一されているものの、詳細な部分になると、項目の階層が異なる場合や、独自の項目名が使用されており、医薬品毎の差異が生じてしまう。

さらに、検索に適していない点も問題である。例えば「胃・十二指腸潰瘍」のように、記述されているものから「胃潰瘍」に関する医薬品を調べても、全文検索で正しい結果が得られない。

また、規格単位が異なる複数の医薬品をまとめて1枚の添付文書に記載しているため、情報が混在している点も問題である。

2.2 医薬品データベースのER図作成

添付文書に書かれている項目を整理するために、「医療用医薬品添付文書の記載要領について」[1]という旧厚生省通知を参考に、基本となる項目を列挙し、ベースとなるER図を作成した。そして、実際の添付文書30種と比較し、不足項目の補完とテーブル構造の改良を行った。販売名のように繰り返しのない項目を「基本データ」としてHOT番号を主キーとしたテーブル(図1)を作成し、そこに禁忌のように繰り返しのある項目のテーブル(図2)を、HOT番号を外部キーとしてそれぞれ作成した。



図1 「基本データ」テーブル(一部)

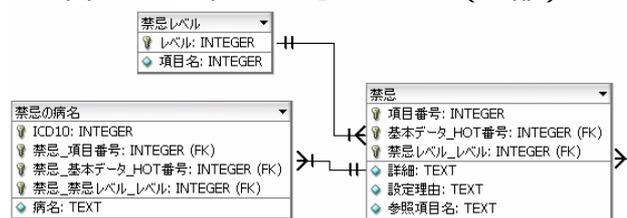


図2 「禁忌」と「禁忌の病名」のテーブル

また禁忌情報に含まれる病名は、図2のように関連させた「禁忌の病名テーブル」にコードとして保持させた。これにより、コードでの検索が可能となり、前出の「胃潰瘍」のような場合の検索が行える。病名コードとしては、標準的に利用されているICD10を採用したが、ICD10が診断名を表すコードであり、記載されている症状を全て網羅できないという問題がある。

2.3 医薬品データ抽出プログラムの作成

PDFファイルからのデータ取得

添付文書がPDF化されたものが、各製薬会社のWebページにて提供されている。これは、一次情報源である紙ベースのものと同じ内容である。よって、まずはPDFから情報を抽出すること目標とし、PDFをテキスト化したものからデータ取得を試みた。アルゴリズムは、項目名を正規表現で表したものにより、データの種別を判断し、取得す

Development of Learning Supporting System for Pharmacists
Keita NABETA[†], Masaomi KIMURA[†],
Michiko OHKURA[†], Fumito TSUCHIYA[‡]
[†]Shibaura Institute of Technology,
[‡]Tokyo Med. & Den. Univ. Dental Hospital

るというものである。結果としては、構造的にまとまっている「禁忌」のような項目は比較的容易に取り出すことができたが、テキストの分離や外字が表示できないなど、PDF がレイアウト重視である点に起因する問題が発生した。そのため、PDF からのデータ取得は困難だという結論に至った。

SGML 形式のファイルからのデータ取得

現在、添付文書の SGML 化は義務付けられており、構造を定義する DTD ファイルに基づいて各製薬会社が作成し、医薬品医療機器総合機構[2]の Web ページにて公開されている。構造が共通化されているため、医薬品毎の基本的な構造の差異はないので、処理の可搬性が広がる。データ取得の流れは以下のようになっている。

() SGML 形式から XML 形式に変換

SGML から処理の容易な XML 形式への変換をする。両者ともタグによってデータを持つという形式のため、XML 宣言の付加や、閉じられていないタグを閉じるなどの処理を行う。この際、BOLD タグのように表示用のタグは除去する。

() DOM ツリーを構築

項目の階層構造を分析し、DOM ツリーを構築することで、処理に適したデータ構造とし、そこからデータ抽出を行う。

() データ格納

取得したデータを、ER 図に基づいて作成した RDB に格納する。この際に、HOT 番号や ICD10 などのコードをマスタから取得し、データに追加する。

以上の処理により、大部分のデータの取得は可能となる。しかしながら、以下のような問題が残る。

(A) item タグと detail タグの利用

タグに項目名、要素にデータという形が本来持つべき姿であるのだが、item タグの要素に項目名、それと並列に存在する detail タグの要素にテキストを格納するという利用の仕方をしている。様々な項目名に対応させることが可能である点では良いが、データへのアクセスが複雑になるという問題がある。

(B) 要素部分の記述に関する問題

データは添付文書の表現をそのまま利用しているため、添付文書の記述方法に関する問題は依然残る。また、英数字の全角半角が医薬品ごとに混在している点や、レイアウトのための無駄なスペースや改行が含まれている点も、データとして扱う際に問題となる。

(C) タグの間違った利用

副作用の項目で、頻度を表す frequency というタグがあるにもかかわらず、item タグに記述さ

れている。また、項目番号を表す serialno タグで同一番号が連続しているなど、正しく SGML 化されていない場合がある。

これらのように、データ側で整備すべき問題点も存在している。

2.4 学習コンテンツの作成

プロトタイプとして、検索システムとテスト形式のコンテンツを作成した。前者は販売名や禁忌の病名をキーワードとして検索するものであり、また後者は、医薬品の添付文書情報を複数提示し学習させ、その後でそれらの医薬品に関するテスト、具体的には禁忌に記述されている病名を選択肢の中から選ばせ、知識定着のチェックを行うものである。(図3)

<p>問題1</p> <p>ジアゼパム錠2mg「アメル」の禁忌の項に記載されている病名はどれか？</p> <p>【選択肢】</p> <ul style="list-style-type: none"><input type="radio"/> 外傷性血気胸・胸腔に達する開放<input type="radio"/> 重症筋無力症<input type="radio"/> 直腸穿孔<input type="radio"/> 突発性難聴
--

図3 コンテンツの一例

3 今後の課題

今後、データベースと学習コンテンツの両方の改良を進める。データベースに関する改良として、現在テキストとしてまとめて取得している用法用量のように、さらに分析してデータに分ける処理の追加を行う。また、外字や上付き文字の様な特殊な表記のデータにはまだ対応できていない。さらに、病名のコード化において、ICD10 にはない症状を表すコードの作成が必要である。例として「身体的疲弊」を表すものは ICD10 にはない。これらの課題を解決し、データベースが整備することで、コンテンツも充実させることができる。現段階では、まだプロトタイプの状態だが、今後はヒヤリハット事例[3]や医薬品の画像データ等、外部の情報とも関連付け、より効果的な学習支援システムを作成する。

参考文献

- [1] 医療用医薬品添付文書の記載要領について
<http://mid.cc.kumamoto-u.ac.jp/606.html>
- [2] 医薬品医療機器総合機構
<http://www.pmda.go.jp/>
- [3] 木村ら “医薬品使用安全性データベースの提案” 日本人間工学会関東支部第 36 回大会
- [4] 中橋ら “SGML による構造化された医薬品添付文書の記述” 第 17 回医療情報学連合大会