

厚生労働科学研究費補助金

医療安全・医療技術評価総合研究事業

標準的電子カルテシステムの導入・普及に関する研究

平成19年度 総括・分担研究報告書

主任研究者 高林克日己

平成20（2008）年 3月

目 次

I. 総括研究報告

標準的電子カルテシステムの導入・普及に関する研究 -----1

高林克日己

II. 分担研究報告

1. スパイラルモデルに基づくクリパスについての研究 -----3

藤田伸輔 木村通男

2. ユニバーサルデータベースについての研究 -----9

藤田伸輔 木村通男

3. テキストマイニングによるカルテ記載からの語句抽出についての研究

鈴木隆弘 -----13

4. 現場で求められている電子カルテのヒューマンインターフェイス

鈴木隆弘 -----19

5. 電子カルテのウイルスに対する標準的セキュリティに関する研究

高林克日己 鈴木隆弘 -----25

平成19年度 厚生労働科学研究費補助金（医療安全・医療技術評価総合研究事業）
標準的電子カルテシステムの導入・普及に関する研究（H18-医療-一般-041）
総括研究報告書

主任研究者 高林克日己（千葉大学医学部附属病医院企画情報部 教授）

研究要旨

電子カルテの標準化といっても現在では単にそのヒューマンインターフェースのみでなく利活用まで多岐にわたる。本研究では昨年に続けて、本年度は電子カルテの開示、クリニカルパスについての研究をさらに進めるとともに、セキュリティ特にコンピュータウイルス対策の標準化、データウェアハウス（DWH）における標準化、カルテ記載に用いられる標準的用語の解析およびユーザーからみたときの基本的標準要件について考えた。電子カルテの開示についてはこの総括研究で詳述する。セキュリティ対策については特にこの一年間に2大学で業務を停止し対策に長時間を要したコンピュータウイルス事件があったことから、現在の国立大学病院のシステムのコンピュータウイルスに対する脆弱性とその標準的対策を自験例と他大学からの調査をもとに述べた。DWHに関しては次第に電子カルテから作成されるようになった DWH の利用について未だ不十分であり、その用途から考えられる標準的仕様を検討した。カルテ記載からのテキストマイニングによる標準的用語の抽出は、それぞれの疾患の中で純医学的な観点から離れて、どのような内容が記載されるべきを経験論的に明示していく重要な作業かと考えられた。またユーザーから見たヒューマンインターフェースの基本的標準仕様は臨床面から最も端的に必要なものは何かを示しているといえる。

A. 研究目的

1) 電子カルテの開示の応用

診療録の開示については裁判所に提出されるような紙のカルテ以上の電子カルテの様式が求められる他、患者中心の医療の時代において、容易に患者が自らの診療情報を入手できる方法さらにはこれが EHR として用いられることも目的とする。

B. 研究方法

患者に開示できる電子媒体としての厚労省版では出力は考慮されていない電子カルテの出力を可能にできるように、千葉大学医学部附属病院において M 言語で格納されている診療記録を HL7ver2.5 CDA(Clinical Data Architecture)に落とし込み、画像データ、検査記録、処方記録と診療情報提供書を合わせた電子診療情報提供書を作成する。

また EHR の先進国である欧州各国の専門家との話し合い、ドイツとの日独医療情報カンファレンス、eHealth07 を通して、欧州における EMR の進捗状況を調査し、その交換規約等について検討した。台湾、韓国における医療情報システムについても現地の病院を視察、聴取した。その他米国における動向も加味しながら、EHR の今後の国際的方向を検討する。

C. 研究結果

1) 昨年からの構想を進めていた CDA への書き出しができるようなシステムを構築し、実用化した。昨年度は厚労省版の電子カルテの作成者から得た知識を基に千葉大学医学部附属病院の情報システムから、CDA を作成する方式を確立し、画像の登録などの手続きを簡便化し、医師の労力を最

小限として、専門担当者によって一定期間の電子カルテからの開示用の CD が作成できるようになった。ただしカルテの表示をどのようにするのか、XML を用いて後利用まで考えた書式を提言したいが現時点で標準的な書式は完成していない。

2) 欧州での視察により、最も EHR が進んでいるとされる英国であっても、病院内ではほとんど電子カルテを用いていない。また N3 と呼ばれる NHS 病院間での医療連携も基盤自身は完成していてもほとんど動いていなかった。ドイツにおいても実験的レベルを超えず、オランダ、北欧などで単純な電文の送受信が行われている程度であった。そもそも院内で電子カルテができていないのであるから、それ以上の interoperability なるものには限界がある。国家間の連携については、画像の伝送や、移住労働者の健康管理など行われているものがあったとしても一般的ではなかった。米国においても RHIO で電子カルテを行っている病院といえども医師が全てを入力しているわけではない。これに対して韓国では急速に電子化が進んでおり、とくに先鋭的な一部の病院における電子カルテシステムはわが国のそれを凌駕している。ディスプレイの2重化、3重化などで一覧性を向上させ、ICUであってさえ全てのモニター情報が自動転送されることでナースの手書きが存在しない施設など、わが国で現在課題とされているほとんどのことが、韓国のこれらの病院では既に解決されてしまっている。レセプトオンライン化もほとんど完成され、すでに EHR としてデータが公開利用されており、この分野においても日本より先を行っている。しかし地域連携としての電子カルテはほとんど行われていない。EMR 先行型という

点ではわが国と同じであった。

D. 考察

1) 商品としての完全な完成品ができあがらないために、実運用でなく、試験運用での確認までしかできなかった。このため今後さらに試験運用を進めながらカルテの開示方法をより良くしつつ、カルテ開示の標準方式を提示する必要がある。また実際にカルテ開示に耐えうるだけの診療録作成を医師側に求めることが極めて重要であることが試験運用中に明らかになってきた。すなわち開示がごく一般的になされるという意識のもとにカルテを書く習慣をまず医師側が持たないとこのようなカルテ開示を行うことは難しいということである。当然のことであるが、これは多数の医師を抱える病院においては簡単に実現することではない。

2) EHR の基本的概念が診療録の一部共通部分の共有であればよいという欧米型に本邦の考え方も全データの interoperability からシフトし、それにともない完全電子カルテの普及の動きが減衰しているように見える。しかしあらゆることにおいて中途半端に行うことが最も非効率的なのはひとつの定理であり、韓国の一部先進病院のように完全なペーパーレス電子カルテができることがすでに不可能ではない時代に、EMR,EHR どちらを向くのかは判断に迷うところである。EMR,EHR のベクトルが同じ方向である韓国との交流はわが国にとっても重要であるし、一方欧米各国では東アジア式の EMR の普及は認識さえしていない人が殆んどで、理解されれば感嘆の的でもあり、決して現在の欧米の方向のみが EMR,EHR の唯一解ではなく、近未来においてこれらすべてが実現される状況が来ると推測される。このような時代においては直近の数年後に実現可能な領域の中で何が最も効率的で、何が公益性を持つかを考えて計画するしかないであろう。現在の厚労省の方針が EHR にシフトして動いているのもこうした一時期の対応であるとの理解のもとで、現実運用されている電子カルテがすでに存在する状況での医療連携、EHR は逆に我が国の独壇場と考えられないことではない。また欧米のある意味では単純化された医療施設間の上下関係に比べて、何でもありの我が国の医療連携におけるインフラとなるべきシステムの構築は、すべての医療施設を物理的には平等につなぐ別形式を要するはずである。今後の日本の医療システムの変革も考えながら構築する必要がある。とりあえずの共有情報はサマリーレベルで十分であるというのが欧米また本研究者を含む日本の多くの研究者で指摘されていることであり、まずこの基本的標準的構造を提唱した上で、それ以上の可能性を実現できる日本の EHR を目指すことが肝要であると思われる。

今回作成した CDA に準拠した電子カルテ開示システムはこのような状況のもと、欧米の EHR 以

上の情報を提示できる国際水準を凌駕したものであるが、カルテの内容を全て自動的に開示することで起こる倫理的な問題など、克服すべき課題はいくつか残っている。患者に直接渡す本方式とは別に、医療施設間のネットワークでの情報共有においても同様のことがいえる。単に一つのサマリーかカルテ全ての情報の開示という、両者の中間レベルで別施設に伝えたい情報を別途作成（あるいは全カルテから選別）することは容易ではないが、全く不可能というわけでもなく、それを行っている地域連携も少ないながら存在する（それらもまたサマリーと呼べなくはない）。そのような方法論は少なくとも残した形で地域情報共有は進むべきであるし、こうしたことで国際水準を凌駕する質の高い EHR を確立する潜在能力を我が国は持っているように考える。

E. 結論

電子カルテの開示の標準書式として提唱されている SS-MIX と電子カルテそのものをセットで CD-ROM にて配布するシステムを構築した。このような仕様は国際的にも高い水準にある。ここでの電子カルテは検索、参照のためであり、これら機能を有する標準化を実現することは物理的に不可能ではないが、倫理的な検討も十分に行うことが必要である。

F 健康危険情報

とくになし

G. 研究発表

1. 論文発表

なし

2. 学会発表

Katsuhiko Takabayashi EHR in Japan.
eHealth 07 (Berlin),2007

H. 知的財産権の出願・登録状況

とくになし

I. 添付資料

なし

スパイラルモデルに基づくクリパスについての研究

分担研究者 藤田 伸輔 千葉大学医学部附属病院地域医療連携部准教授
分担研究者 木村 通男 浜松医科大学医療情報部教授

研究要旨

電子カルテにおいて患者安全の確保と診療の質を担保する手段として、また医療者の負担を減じる手段として電子クリパスの活用が重視されている。本研究では標準治療計画にそった症例数の増加という視点ではなく、標準治療計画から外れる症例の早期発見とその対処を重視したスパイラルモデルの実証を目指した。千葉大学医学部附属病院の電子カルテが平成19年よりVersion 4へとシステムアップした際に電子クリパスにスパイラルモデルを導入したので報告した。

A. 研究目的

クリパスの導入目的は①標準治療の実施、②効率的な医療の実施、③患者安全の推進、④異常の早期発見、⑤わかりやすい医療の提供(患者と医療者の共通理解)などが挙げられる。しかしこれまではパスの適用率を重視するあまり①および②が重視されすぎた感もある。本研究では③、④、⑤を重視したクリパスのあり方の検討を目指した。

B. 研究方法

本研究では以下のテーマの基づき検討し、医療機関情報システムでの実用を目指した。

1. クリパスの目的、使われ方の検討
2. 紙のクリパスと電子クリパスとの比較
3. スパイラルモデルに基づくクリパスの検討
4. 地域連携型パスへの応用の検討

(倫理面への配慮)

本研究においてはクリパス構築法についての研究であるため倫理的問題は発生していない。

C. 研究結果

1. クリパスの目的、使われ方の検討
「クリニカル・パス」もしくは「クリティカル・パス」で検索された論文をもとにクリパスの目的と使われ方を検討した。
 - 1.1. 標準治療の実施
クリパスの主たる目的として標準治療の実施が挙げられている。標準治療としてガイドラインが示されている疾病・治療法もあるが、定められていないものも多い。クリパスの中で扱われている標準治療はガイドラインに示された治療のみならず、各施設内で合意に達した治療法という意味でも使われている。ガイドラインの電子化規約としてGLIFがあるが、これとクリパスとを連携させる試みは未だなされていないようである。またパスの適用率、バリエーション率をもって標準治療の達成率とみなすことも可能であるが、クリパスの成果を標準治療の改善などに活用する手法について

は述べられていない。

- 1.2. 効率的な医療の実施
EBMに基づく医学的な標準治療、EBMには至っていないが、施設の中で妥当と考える標準治療という考え方とは別に、診療を効率的に行うためのクリパスという考え方も存在する。効率性は検査機器などの運用効率を考慮したもの、入院期間の短縮など時間的効率を考慮したもの、医療者の作業効率を考慮したもの、診療報酬制度に基づく経済性を考慮したものなどがある。これらについてもクリパス運用の成果を還元している報告はまれである。
 - 1.3. 患者安全の推進
クリパスの中で患者安全増進のための手順を盛り込んだものがある。また術前検査など幅広い意味では患者安全の推進としてカウントすべきかもしれないが、「患者安全の推進」を謳いながらクリパスの中でどのように患者安全が推進されたかは不明である。
 - 1.4. 異常の早期発見
クリパスから外れたバリエーション症例が何らかの問題を抱える症例であることは容易に想像できる。従ってクリパス適用患者の中から異常を生じた異常を生じた患者を抽出することも可能である。
 - 1.5. わかりやすい医療の提供
治療計画を明示することはチーム医療においても、あるいは治療対象となる患者とその家族にとっても有用である。
 - 1.5.1. チーム医療
多職種によるチーム医療ではチーム内での共通理解のため日程を含めた具体的な診療計画が必須であり、クリパスはチーム医療の推進に有用である。
 - 1.5.2. 患者理解
患者自身が自己の治療過程を理解することはインフォームドコンセントという面ばかりでなく、治療効果を高めるためにも重要である。文章で書かれた説明書きよりも日程

管理された治療スケジュールの方が理解しやすく、クリパスは有用である。ただし患者理解のためのクリパスとチーム医療のためのクリパスとでは書かれるべき内容も表現も異なるべきである。

2. 電子クリパスの現状

クリパスを扱うプログラムは現在大きく次の3段階に分類できる。電子クリパスでは紙に書かれたクリパスの情報をオーダプログラムとして実現することに主眼が置かれている。

2.1. 紙ベースのクリパス作成支援プログラム

標準治療の実施、効率的医療の実施、患者安全の推進、わかりやすい医療の提供などクリパスの機能として述べられていることのほとんどを紙ベースのクリパスでも達成しうる。すなわち現在も紙ベースのクリパスの意義は薄れていない。そこで紙ベースのクリパスを作成するプログラムは現在でも有用である。

2.2. オーダ発行を主目的とするオーダ発行クリパスプログラム

電子カルテに実装されているクリパスプログラムではクリパスを個人に適用する際に半自動的にオーダ発行するものが多い。半自動オーダ発行とは祭日への対応や入院の曜日による差を処理するために必要な機能である。このためにオーダプログラムと密接に連携しており電子クリパスの開発性を悪くしている。

オーダプログラムと密接し、診療行為とは連携していないためクリパスは線形に構成される。勢い経過が多彩な大きな手術や内科慢性疾患への対応は困難となっている。

2.3. 臨床経過制御型クリパスプログラム

電子カルテにおいて診療行為をモニターし、診療過程そのものをコントロールしようという試みがなされている。ループ処理、条件分岐、エラー処理などを扱えるためほとんどの症例をバリエーションとしないですむ。しかしクリパスは電子カルテと密着するというよりもむしろクリパスのもとに電子カルテを構築する形態となるため、クリパスのバージョンアップと電子カルテのバージョンアップとを連携させる必要がある。これら2つのプログラムをマルチベンダー化することは困難である。

3. スパイラルモデルに基づくクリパス

スパイラルモデルに基づくクリパスでは臨床経過制御型とすることも可能であるが、紙ベースのクリパスでもオーダ発行を行うクリパスでも対応可能とする。これらすべての形態に対応し、マルチベンダーに対応するためにクリパスクリエータ、エディタ、オーダ、レポート、アナライザに分割し(図1)、各プログラム間の通信を標準化する。これらのモジュールはすべて独立している必要はなく、いくつかの機能を一つのアプリケーションで提供しても問題はない。ただ、診療の電子化に従って各モジュールのインターフェースおよび要件は異なるためモジュール単位に独立したプログラムとして以下にその要件を述べる。

3.1. クリパスクリエータ

標準クリパスを作成する。クリパスの名前、適応疾患、ステージ、作成者、作成年月日、作成支援者などの情報からなるヘッダ、クリ

パスそのものとなるアクト集、標準的打ち切りポイントを設定し、作成したクリパスを標準形式で出力する。

標準クリパスの中で繰り返し処理、条件分岐、ジャンプ処理をサポートしなければならない。またすでに作成したクリパスをサブルーチンとして実施可能でなければならない。

3.2. クリパスエディタ

クリパスエディタでは標準クリパスを読み込み、症例に適應するように編集を加える。また合併症に対応するため複数の疾患の標準クリパスを読み込み、これらを一本にまとめて個人用クリパスを作成する。この際、クリパスから逸脱するポイントを作成し、クリパス適用を中止する場合の理由と逸脱したポイントを明確にする機能を有しなければならない。

診療の過程において新たな診断が追加されたり主たる診断が変更されたりすることもある。こう言った場合には現在運用中のクリパスに対して変更をかけて運用するのではなく、クリパスレポートによって運用中止の理由を記載し一旦クリパスの運用を中止する。その後エディタを用いて新たなクリパスを作成しなければならない。

3.3. クリパスオーダ

クリパスエディタで作成した個人用クリパスに従ってオーダを発行する。オーダは全過程を一気に発行するのではなく、分岐ポイントまでとし、分岐ポイントにおいて継続の判断、あるいは違う選択肢を選択することを許容し、次のポイントまでのオーダを発行する。この方針により電子カルテプログラムとの過度な依存関係を回避する。

3.4. クリパスレポート

個人用クリパスに則って診療を行い、診療行為の実施記録およびその結果を記載する。また診療過程で下された診断およびパスからの逸脱および完遂を記録する。

3.5. クリパスアナライザ

クリパスレポートの結果を集計し、分析する。アナライザでは必ずしもクリパス専用のプログラムである必要はなく、分析者のスキルに応じてマイクロソフトエクセルなど汎用のプログラムや各種の統計プログラムや様々なツールを用いても全く支障はない。

クリパスクリエータよりレポートまでの書式を統一することにより多施設間でのクリパス運用結果解析を可能とできることが重要である。

D. 考察

診療過程の標準化を目指しクリティカルパス、クリティカル・パス、クリティカルパスウェイ、クリティカル・パスウェイ、クリニカルパス、クリニカル・パス、クリニカルパスウェイ、クリニカル・パスウェイ、クリパスなど様々な表現が用いられている。これらの表現はそれぞれ使用者の思いを表現した用語であり、統一することは現時点では不可能である。本稿ではこれらの立場について考慮しつつも取り扱いの簡略化の観点からすべての表現を略称したのものとして「クリパス」という表現を使用した。

クリパスが目指す標準治療の実践が患者安全に寄与し、かつ医療費の削減にも寄与することは異論のないものとする。しかし我々がMedlineなどを用いて調査した限り、現在各診療機関で行われている診療が最善のものであるか否かを分析することを強調した報告は発見できなかった。標準治療は医学の進歩に伴い変遷するものであり、何が問題でどこを工夫したからどのような効果が上がったかを記載し、新たな標準治療をできることが医療において重要であり、一医療機関内での解析ではなく複数医療機関での解析が可能でなければならない。

例えば潰瘍性大腸炎に対する治療では、罹患大腸をすべて摘出すれば愁訴を中心とした問題は解決するはずだと考えて大腸全摘が試みられた。しかし生活の質を考えて肛門機能を温存しようとして様々な試みがなされるも惨憺たる結果を経て肛門温存手術は禁忌とさえいわれた時代があった。それでもなおQOLを改善したいという外科医の熱意はJ-pouchやS-pouchといった術式を生んだ。我々は実際にこのような変遷をJ-pouchを開発した宇都宮の指導のもと経験し、その手術結果を分析して術式の改良に励み非常に高い患者満足度と疾病の良好なコントロールを維持しながら手術回数を3回から1回へ、手術合計時間を12時間から2時間強へと改善してきた。このような診療の改善を客観的なデータを示しながら行っていく試みの中で現時点での標準治療を記載し、その治療を実践していく上での問題点を正確に記載し、結果を分析して改善すべき点を明らかにするというプロセスの重要性を痛感した。本研究はこのような診療プロセス改善のあるべき姿を求めた研究である。

本研究では診療行為は日々同じような行為を繰り返しながら絶えず変化していくという点で日々の繰り返すと少しの前進という意味でのスパイラル、および計画の立案・実践・結果・分析・改善というPDAサイクルを回すという意味でのスパイラルを合わせてクリパスのスパイラルモデルと命名した。

自己の診療を振り返り、問題点を発見し、改善策を探り、新たな診療方針を実践した結果をもとにさらなる改善を図るという医学本来の基本方針を重視すれば、診断・治療・評価の標準化の重要性に異論はないはずである。本研究ではこのような観点の下にクリパスの意義を再検討し、クリパスのあるべき姿を検討し、クリパスクリエイター・エディター・オーダ・レポート・アナライザという機能分担に至った。

クリパスの中で使用する用語としてはMERIT-9を標準とし、MEDIS標準規格週に沿ったものとする。クリパスの構造として図2、図3をしめした。

E. 結論

クリパスの意義を再考し、クリパスがPDAサイクルを通して成長するという意味を込めて、また日常診療が日々の繰り返しの中で進められて行くためにクリパスもこの螺旋を踏襲すべきという意味を込めてスパイラルモデルと名付けた。多施設間でのPDAサイクルを通じた比較を容易にするために標準規格を策定し、その上にスパイラルモデルを構築する基本を示した。

F. 健康危険情報

本研究において健康危険事項は発生しなかった。

G. 研究発表

1. 論文発表
準備中

2. 学会発表
準備中

(発表誌名巻号・頁・発行年等も記入)

H. 知的財産権の出願・登録状況(予定を含む。)

1. 特許取得
ありません

2. 実用新案登録
ありません

3. その他
ありません

図1. クリパスツールの構成

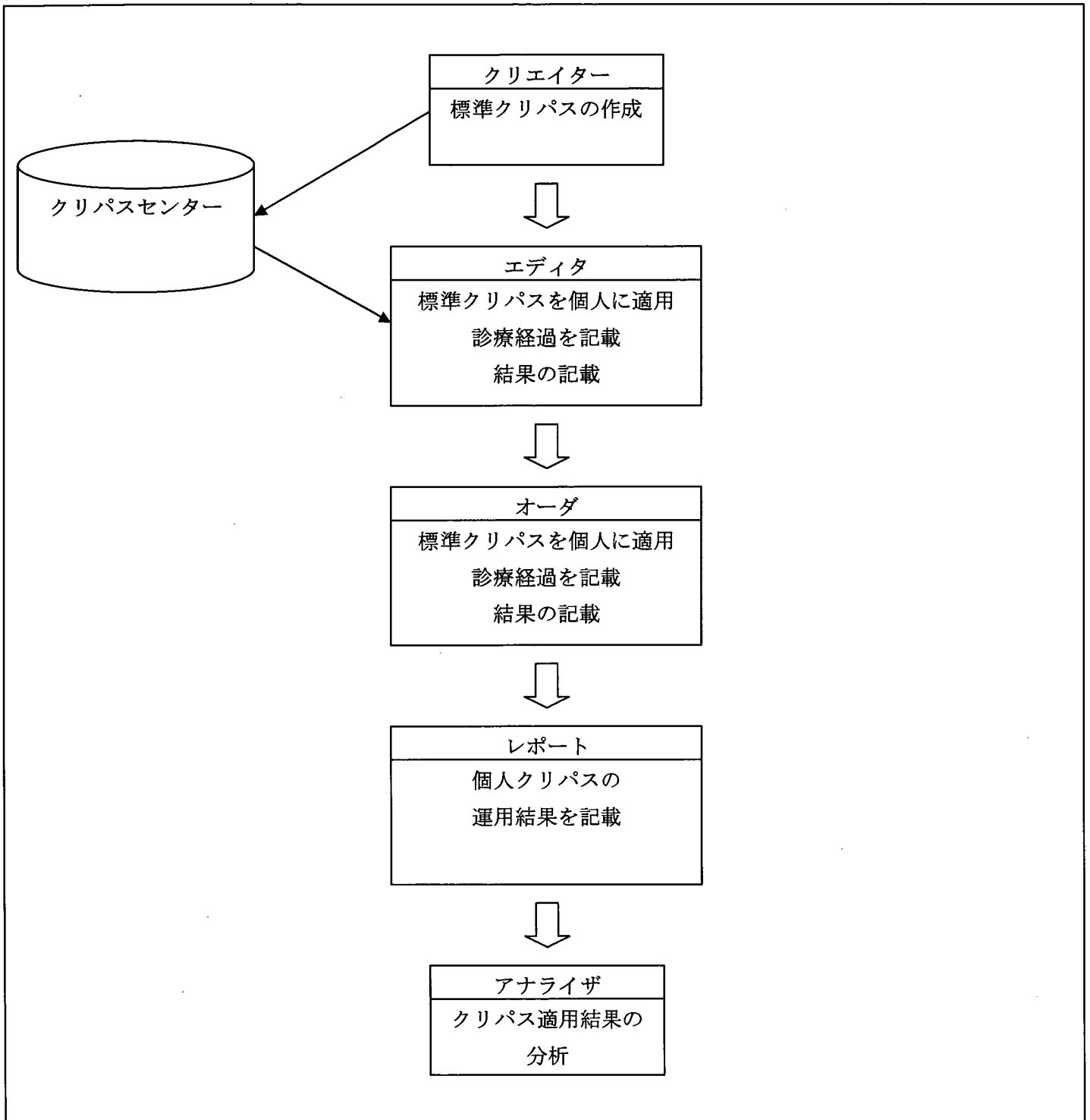


図2. スパイラルモデル標準クリパスの構造

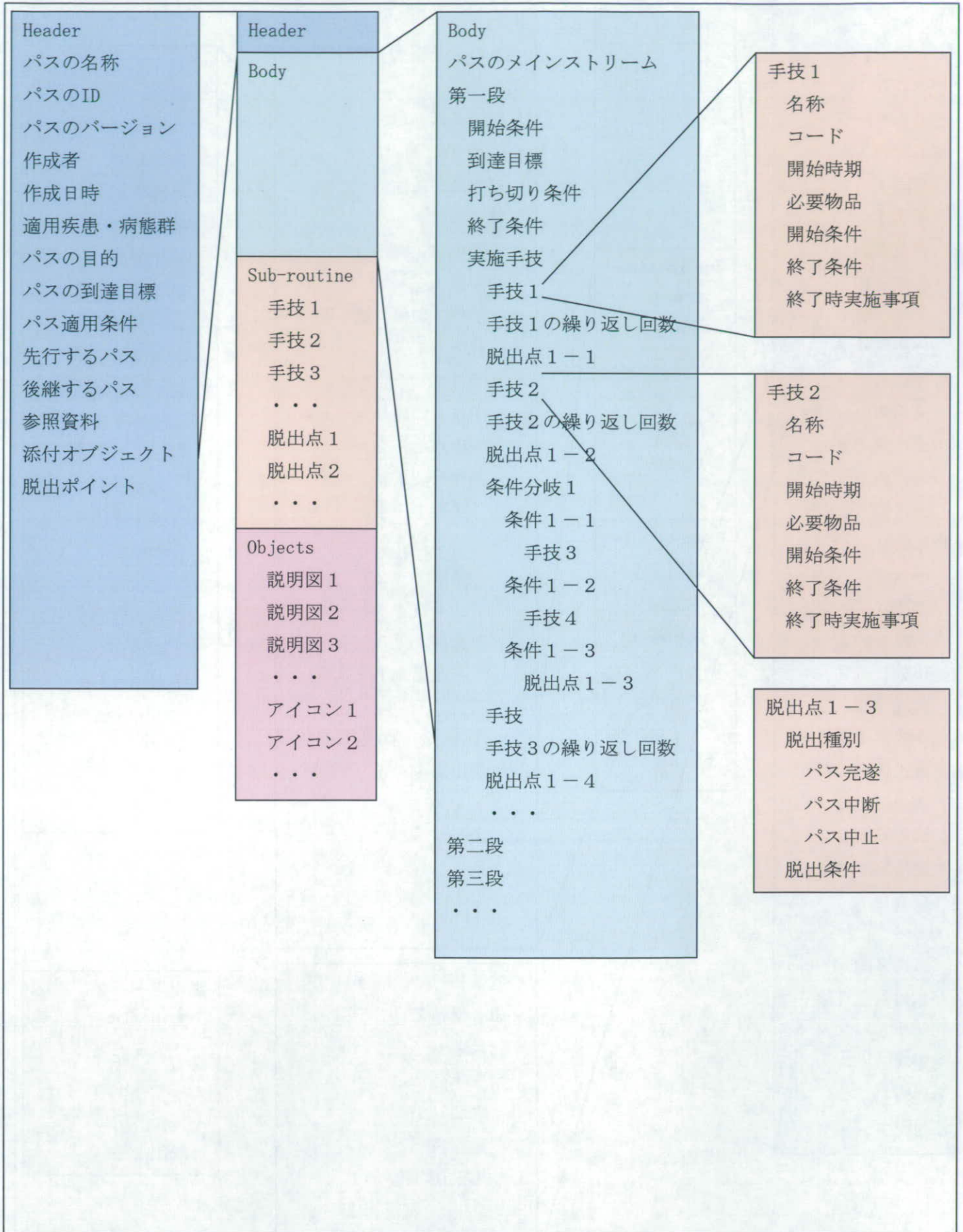
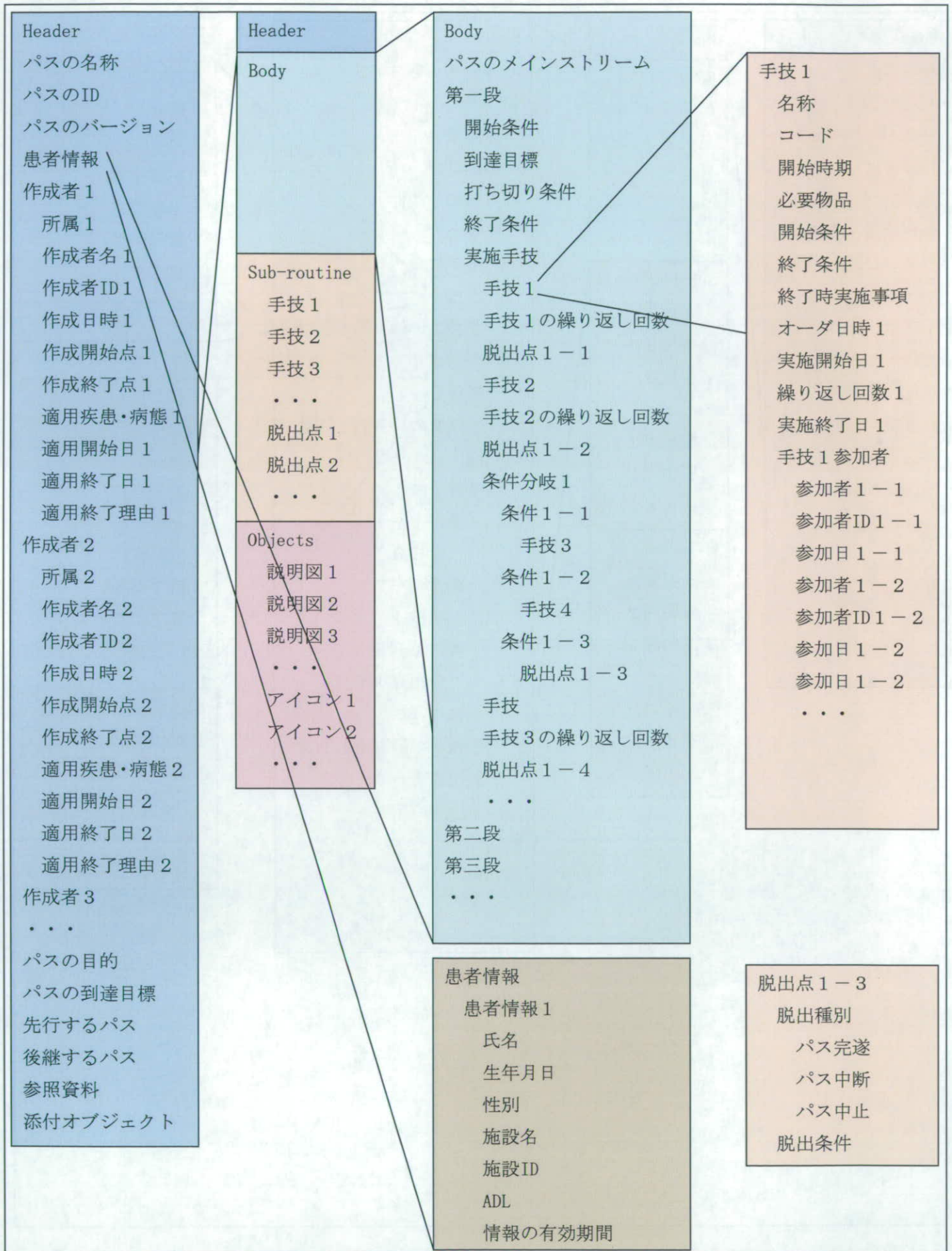


図3 スパイラルモデル運用クリパスの構造



ユニバーサルデータベースについての研究

分担研究者 藤田 伸輔 千葉大学医学部附属病院地域医療連携部准教授
分担研究者 木村 通男 浜松医科大学医療情報部教授

研究要旨

電子カルテの有効利用として診療データを分析し、診療に役立てるデータウェアハウス (DWH) として利用することがある。本研究ではDWHへの問い合わせ言語の標準化と分析手法について検討した。問い合わせ方法を一般化するためにHL7のQueryメッセージを基本とした。またデータ分析方法としてテキストマイニングについて検討し、その有効性を確認した。

A. 研究目的

電子カルテのデータをデータウェアハウス (DWH) として活用して診療の改善、臨床研究の推進、病院機能の改善などを推進することが期待されている。本研究ではDWHと活用する上で必要な手法を開発することが目的である。

DWHとして活用する際に多施設で同一手法が使える場合も多いと思われ、あるいは多施設間共同研究も考えられる。このような用途では問い合わせ応答速度よりもデータベース構造に影響されない問い合わせ手法が望まれる。一般的なデータベース問い合わせ言語としてSQLがある。SQLではデータベースに対してアクセス権を有し、データベース構造を理解していることが前提となる。医療分野でのDWH利用においてはアクセス権管理が重大な問題となるため、アクセス権管理を問い合わせ言語と分離すること、データベース構造を公開しないことが必要と考える。そこで本研究ではDWHへの問い合わせを中間言語化し、DWH内でこの言語を翻訳して問い合わせる手法について検討した。

臨床データをDWHとして利用する際には時系列に従って得られたデータを縦横に分析することが必要となる。たとえばコホート研究から得られた定期的繰り返し観測データではデータセットを立方体のようにとらえて縦横断面で切り出すパネルデータ分析の手法が有用である。しかし一般に臨床データ分析では健康状態が悪化する際には密に多項目のデータが収集され、回復すると観察項目も頻度も減少する。このような不規則なデータセットではパネルデータ分析手法は適用困難である。そこで本研究では観察が時間的にも项目的にも密になることを利用した解析手法としてテキストマイニング手法を適用することを検討した。

B. 研究方法

1) ユニバーサルデータベース問い合わせ

病院情報システムをマルチベンダーで構成する際にデータ交換をユニバーサルシステムとして実施するためにHL7が存在する。そしてHL7のVer.2シリーズにはデータ問い合わせ用にQueryメッセージが用意されている。すなわち患者情報を問い合わせる際には、A19を用いて問い合わせメッセージを構

成する。

QRY^A19^QRY A19

QRY：他システムへの問い合わせ

QRD：問い合わせ定義

QRF：問い合わせフィルター

これに対して患者情報データベース側は返答メッセージを以下の要素を用いて構成する。

ADR^A19^ADR A19

ADR：患者管理システムからの応答

ADT：

PID：患者番号

PV1：患者来院情報

...

このような問い合わせにおいて患者情報データベースへのアクセス権確立は問合せの前に実施され、データベースの構造に対する知識は不要である。すなわち本研究の前提条件を満たすものである。そこでHL7 Ver. 2.5を用いて医療用ユニバーサルデータベースにおける問い合わせ条件を検討した。

2) テキストマイニング解析

診療経過において必要なタイミングで実施される検査データは検査項目も測定間隔も全く不均一である。また検査項目の中には異常を生じていないことを確認する目的で実施するものや正常値であることが鑑別上有用なために検査する項目も含まれている。さらに臨床検査データの正常値は同一医療機関内であっても試薬の変更や測定機の変更、あるいは測定方法の変更などによって正常範囲が変化するため、数年以上の期間にわたって収集されたデータを数値データとして解析することは問題がある。そこで本研究では異常低値、正常値、異常高値の3種のみ分類し、例えばASTでは"AST-L", "AST-N", "AST-H"の3表現で置き換え、これらを一単語とみなした。さらに個人のデータを日付ごとにまとめたものを一パラグラフとみなし、個人データ全体を一篇文章とみなしてTF-IDFによって解析を行った。またダッシュ以下を無視して、すなわち-L, -N, -Hを無視した解析と-L, -N, -Hのみを抽出した解析を行った。

(倫理面への配慮)

本研究の成果を実際の医療用DWHに適用する際にはアクセス権の処理および問い合わせに対する返答時に必要な匿名化の範囲を決定する必要がある。しかし本研究においてはこれらアクセス権と匿名化の検討は対象範囲外とした。また千葉大学医学部附属病院のDWHにおける検証では一般ユーザの使用できない環境において実施し、設定条件に対して得られた結果はただちに削除したうえでFILE SHREDDER (<http://www.files shredder.org>)を用いて痕跡まで完全に削除したため倫理的問題は発生していない。

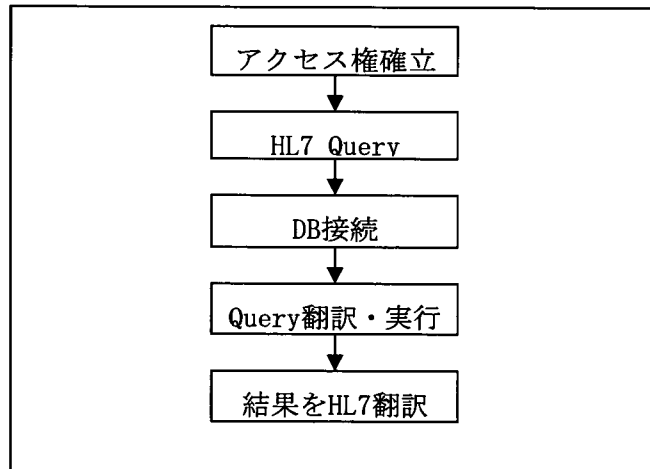
C. 研究結果

1) ユニバーサルデータベース問い合わせ

以下にHL7 Ver2.5におけるQRYメッセージの一覧を列挙する。

A19	QRY/ADR-患者照会
P04	QRY/DSP-請求書とA/R明細書の生成
PC4	QRY-PC/ プロblemsの照会
PC9	QRY-PC/目標の照会
PCE	QRY-PC/診療計画(Problems指向)の照会
PCK	QRY-PC/診療計画(目標指向)の照会
Q01	QRY/DSR-即時応答用に送信された照会
Q02	QRY/QCK-遅延応答用に送信された照会
R02	QRY-検査結果の照会
R03	QRY/DSR 表示指向の結果, 照会/非要求の更新(下位互換用) (Q05に変更)
T12	QRY/DOC-コメントの照会

データベースの構造を知らなくてもQRYメッセージはこのようにさまざまなデータベースに接続し、QRD、QRFコマンドによって詳細なQueryを発することが可能である。例えば血清生化学検査結果から異常高値を呈したもののだけを抜き出すといったことも可能である。



2) テキストマイニング解析

血清生化学検査の結果を用いて検討を加えた。

サンプルデータは千葉大学医学部附属病院の病院情報システムに平成19年1月1日から12月31日までに登録されたものである。検査データに付加された「低値」、「正常値」、「高値」の判定を基に検査名と合わせて"AST-L", "AST-N", "AST-H"という表現を作成した。

患者番号の匿名化操作として、あらかじめ乱数によって並べた匿名化用番号テーブルを作成し、サン

プルとなった患者番号を若い順に並べたテーブルと対応させて変換した。なお氏名・性別・生年月日等のデータは削除した。作成したデータのサンプルを示す。

3131713, 20070108, AST-N, ALT-N, rGTP-H, UN-H, CRN-H
3131713, 20070122, AST-N, ALT-N, rGTP-N, UN-N, CRN-H

これは同一人物に対して2007年1月8日と22日に行った検査結果である。同一日に2回以上の検査を行った場合、2行の検査結果を作成した。テキストマイニング解析では"AST-N", "ALT-N"などを一語として扱い(検査結果統合解析)、一回の検査を一文と、一人分の全データを一篇文章としてTF-IDFによる単語の出現頻度解析を行った。さらにハイフン以降を無視した検査名のみによる検査名出現頻度解析、およびハイフン以降のみを抽出した検査結果単独出現頻度解析を行い、さらにクラスタ分析を行って診断との一致を見た。診断の分類にはICD-10およびDPCのMDCを用いて検討した。

検査名出現頻度解析では同一MDCの中からDPC下位コード分類が可能であった。これは検査結果統合解析よりも分解能に優れているようであった。一方検査結果単独出現頻度解析は有用と言えなかった。

D. 考察

HL7は医療情報システムに特化した情報伝達規約であり、患者情報管理システム・入院管理システム・検査システム・画像システム・給食システムなどある程度定型のシステム構成であることを前提としているが、サーバ構成やディレクトリ構成、あるいはデータベースの構造などに規定されずに使用できる。その開発意図はマルチベンダーによる医療情報システム構築を目指したものである。マルチベンダーを前提として様々な医療機関で使用可能とするためにデータベース構造に依存しない情報伝達規約を作成したが、当然の帰結としてユニバーサル化に伴うオーバーヘッドの増大を招き、システム性能が低いとレスポンスが大きく低下する。我が国は世界の中でも早期から電子カルテ化に取り組んできたため、低いCPUパワー、少ないディスク容量でのシステム構築を余儀なくされたことが大きな要因となっている。オーダーシステムから電子化に着手して30年余の歴史を持つ千葉大学でもHL7への対応は長く見送られてきた。しかし医療機関の機能分担が進み多施設間連携が大きなテーマとなり、電子診療情報提供書、患者向け電子診療情報提供書、特定健診結果などがHL7 CDAをベースとしたSS-MIXされた。このSS-MIXが普及するに伴い各医療機関のHL7対応が進むものと期待される。このような判断から独自のユニバーサルデータベース問い合わせ言語を作成するよりもできる限りHL7に準拠することがよいと考えた。またHL7に準拠することで海外の医療機関でも対応可能となる可能性が高いことを強調したい。

各医療機関の情報システムで蓄えられた診療情報を患者個人の治療以外に活用するという考え方は西洋医学の根本であるともいえる。医療者の興味を引いた病態について症例報告を作成し、同じような病態をとった症例をまとめて臨床研究報告を行い、さらにこれを基礎医学の立場から説明することによって医学が発展してきた。世界中の医療者が共通の評価軸に立ってその研究成果を競い合うために学会活動を行ってきた。そのために国際疾病分類に始まり

医療情報の標準化が進められてきた。その結果多施設の診療記録を横断的に、また時系列によって縦断的に分析すればもっといろいろなことが分かるのではないか、より安全で効率的な医療を抽出し、人類の健康に貢献できるのではないかという期待は高い。米国病理学会よりSNOMED-CTを購入したIHTSDO諸国の狙いもまさにこの点にあると思われる。一方、より安全で効率的な医療を商用に活用し医療・健康関連産業において自社の成績を伸ばしたいという要求も非常に高い。このような利用法に対してどのようにコントロールしていくべきかという議論は学術と商用の境界が不明確であること、商用であっても人類の健康増進に貢献しているという側面も否定できないことから、高度に政治的判断を要する。従って医療分野のユニバーサルデータベースを作成することの是非については本研究の範疇を超える論議が行われることを期待する。

近年新興感染症が問題となり、International Health Regulation が大きく改定されてIHR2005となった。この中では国はWHOから新興感染症に対するモニター体制の整備と迅速な報告義務が課せられている。またC型肝炎問題など医療安全に対する国民の関心の高まりから迅速な情報収集体制の確立が求められている。医療分野のユニバーサルデータベースはこれらの分野に対しても貢献できる技術である。すなわちユニバーサルデータベースを作成することの是非とは別にそれを構築する技術を確立し、その活用方法について論議を深めなければならないと考え本研究を遂行した。

医療分野のデータは匿名化されているか否かにかかわらずその取扱いに十分な配慮をなすべきであり、外部から直接データベースにアクセスできるべきではない。統計調査を含めて国からの紙ベースの問い合わせはデータに対するアクセスを明確にできるという長所を持っており、この調書を保ちながらデータ抽出の時間と労力を削減することが必要である。そこで本研究では問い合わせ言語を統一し、調査を希望する者はこれを医療機関の管理者に提出し、医療機関の管理者の判断によってデータベースへの問い合わせを行い、その結果を監査したうえで調査依頼者に回答する形式(図1)を採用した。

医療分野のユニバーサルデータベース作成に、あるいはその活用にどのような問い合わせがなされるのかについて本研究で想定したものは、厚生労働省からの統計調査、IHR2005対象患者の抽出、薬害などの発生時に対象者の抽出、治験症例のデータ抽出である。統計調査では特定期間の患者抽出をベースに診療情報を抽出できればよい。このため患者来院情報を抽出後、得られた患者リストに対して順に必要な情報を取得する問い合わせはHL7のQueryでカバーできる。IHR2005への対応では病名から対象患者を抽出すればよいためHL7で対応できる。薬害等での抽出では特定薬剤や医療機器・材料の使用をHL7で抽出できる。治験では対象患者リストをもとにその診療情報をHL7で抽出できる。すなわち今回想定した用途ではHL7標準のQueryですべて対応可能と判断した。HL7標準では"AND"や"OR"を駆使した複合的問い合わせには対応困難な事例の発生も考えられるが、検索の掛け合わせ手順でほとんど対応可能と思われる。

データ解析については本研究で検査結果をテキスト化してTF-IDFを用いる方法を示した。多施設のデータを扱う際に、あるいは同一施設であってもある

程度長期間のデータを扱う際に正常値の範囲が一定でないことが解析を困難にする。この問題は検査オーダーシステム導入時より問題となり、標準偏差を用いて補正する方法、正常範囲のずれに合わせて検査値を補正する方法など様々な提案がなされてきた。本研究では検査結果が正常範囲にあるのか、外れているのかという3値化による統計処理を提案した。この変換によってどの程度以上なのか、変化速度は大きいのか、といった情報は失われ、また統計処理方法にも制約を受ける。しかし長期間の追跡研究で区間中のどのデータを用いて解析すべきなのか、あるいは区間中のデータを平均すべきなのかといった問題を回避して得られたデータはすべて用い、測定していない項目、正常値を確認した項目、合併症を有するために測定した項目などの影響をテキストマイニングによって分析できる可能性を見出した。

本研究においてこのようなアプローチをとった理由の一つに慢性疾患の状態判断がある。先進諸国のみならず開発途上国においても慢性疾患のコントロールは大きな課題となっている。

E. 結論

医療分野のユニバーサルデータベース構築とその利用についてHL7のQueryによる問い合わせでデータ抽出できることを示した。また抽出結果の分析手法としてテキストマイニングが応用できることを示した。

F. 健康危険情報

本研究において健康危険事項は発生しなかった。

G. 研究発表

1. 論文発表
準備中

2. 学会発表
準備中

(発表誌名巻号・頁・発行年等も記入)

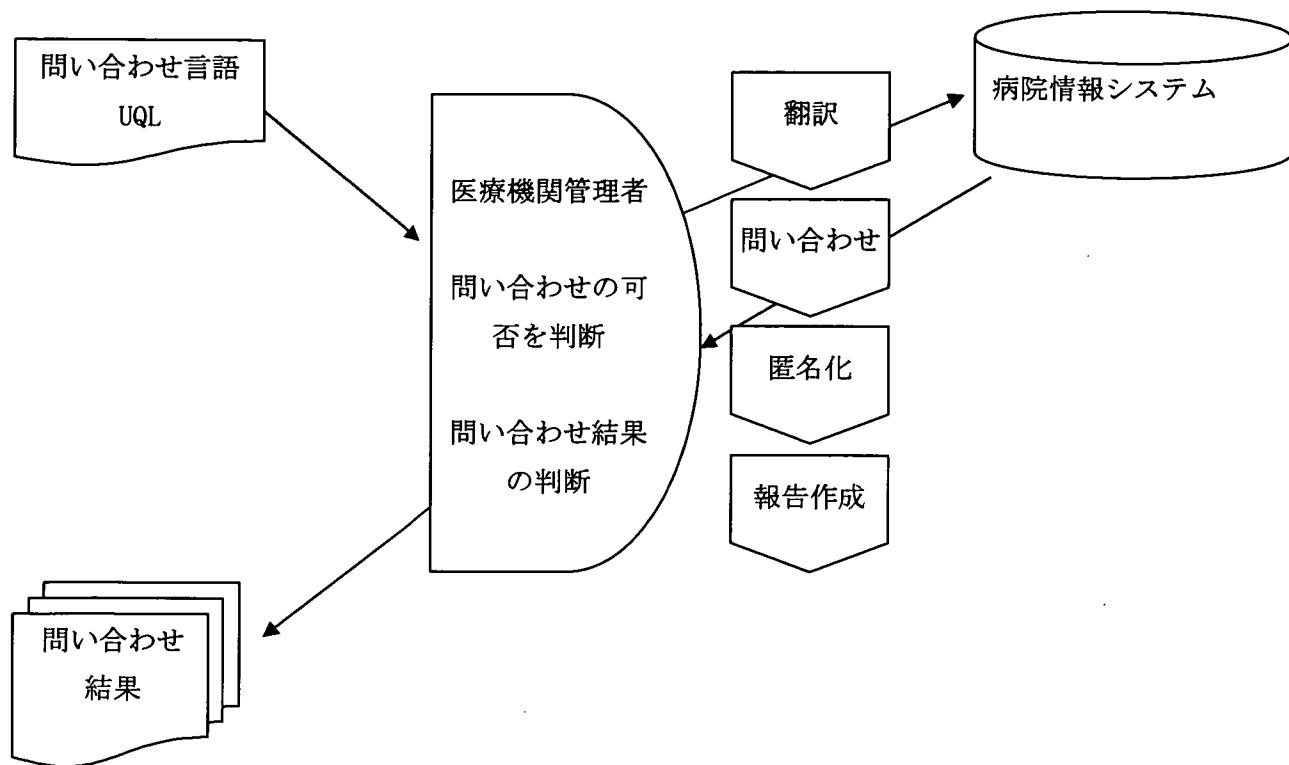
H. 知的財産権の出願・登録状況(予定を含む。)

1. 特許取得
ありません

2. 実用新案登録
ありません

3. その他
ありません

図1. 問い合わせ手続きの模式図



テキストマイニングによるカルテ記載からの語句抽出についての研究

分担研究者 鈴木 隆弘 千葉大学医学部附属病院企画情報部講師

研究要旨

DPC (Diagnosis Procedure Combination) の問題点として、レセプトでの診療内容のチェックが難しくなることが指摘されている。我々はDPC毎にサマリーを解析して疾患に共通、あるいは特徴的な単語を抽出すると共に、DPCを判定することを試みた。

千葉大学医学部附属病院情報システムに保存された平成16,17年度の約17000件の退院サマリーとDPC情報を解析の対象とし、このうち症例数が20例以上の126種類のDPCを対象とした解析を行った。形態素解析にはMeCabを、医学専門用語辞書としてはPHYXAM用語集を採用し、得られた索引語の重要度はTF×IDF法を用いて求めて文書ベクトルを算出した。

索引語抽出では、約1万4千のDPCを特徴づける、あるいは疾患に共通してよく使用される単語が抽出された。DPCの判定では、80%と高率にDPCを特定することが可能であった。

A. 研究目的

診療報酬の包括評価であるDPC(Diagnosis Procedure Combination)が特定機能病院を対象として開始されてから2年が経過し、当院においてもデータが蓄積されつつある。DPCの問題点として、レセプトでの診療内容のチェックが難しくなることは導入前から指摘されており、導入後の臨床の現場からは、レセプトの作成や監査においてどのDPCを選択するべきか判断に迷うことも少なくないという意見がよせられている。我々はベクトル空間モデルを応用することで、疾患を特徴づける、あるいは疾患に共通する索引語の抽出や退院サマリーからのDPC自動判定を試みた。

B. 研究方法

1) 対象

千葉大学医学部附属病院情報システムに保存された退院サマリー文書のうち平成16年度以降の約17000件の退院サマリーとDPC情報を解析の対象とした。このうち16年度退院分の約11000件をDPC別の文書ベクトル空間の構築に用い、17年度退院分の情報をDPC自動判定の検証に用いた。(図1)

症例数20例以上の126種類のDPCを対象とした解析を行い、ベクトル空間の構築には5927症例、自動判定の検証には1988症例が選択された。(表1)

2) 方法

1. 形態素解析と専門用語辞書

自然文の解析には、文字列を名詞や形容詞、助詞といった要素に分解する形態素解析と呼ばれる行程が必要である。本研究では、形態素解析器として奈良先端科学技術大学自然言語処理学講座の開発した「MeCab(和布蕪)」を使用した。医学専門用語辞書としてはPHYXAM用語集を採用し、病名、薬剤名、検査名の辞書は各々当院のマスターから作成して追加した。辞書に存在しない略語や単語を反映させるために未知語からキーワードとして適当と思われる単語を抽出して辞書に

追加した。更に、形態素解析後のテキストをチェックして、適切な切り分けがされていない場合にも追加登録を行った。

2. ベクトル空間モデル

対象とする退院サマリー集合全体をDとし、DPC毎の退院サマリーを $d_1, d_2, \dots, d_j, \dots, d_n$ とする。またDから抽出したm語の索引語を $w_1, w_2, \dots, w_i, \dots, w_m$ とする。ここで、退院サマリー d_j における索引語 w_i に対する重みを a_{ij} とおく。このとき退院サマリー d_j はm個の要素 a によってベクトル表現が可能となり、これを疾患jの退院サマリベクトルと定義した。(図2)

3. 重み付け

本研究では索引語の重み付けにTF/IDF法を用いた。

(図3)

重み α_{ij} は $\alpha_{ij}=(l_{ij} \times g_i)/n_j$ で表す。

$l_{ij}=\log(1+f_{ij})$: l_{ij} は局所的重みであり、索引語 w_i の文書 D_j における出現頻度に基づき計算され、文書中に頻繁に出現する索引語に対して大きな値が与えられる。今回は索引語 w_i のDPC分類jの退院サマリー d_j における出現頻度 f_{ij} を用いた。

$g_i=\log(n/n_i)$: g_i は大域的重みであり、文書集合全体に亘る索引語 w_i の分布に基づいて計算され、特定の文書に集中して出現する索引語に対して大きな値が与えられる。 n は退院サマリ-の総数であり、 n_i は索引語 w_i を含む退院サマリ-数である。

n_j : 退院サマリ-の長さによる重み付けの影響をなくす文書正規化係数である。

4. 類似度の判定

2つの退院サマリベクトル間の類似度はベクトルの内積 $d_j \cdot d_k$ によって定義した。もともと類似度が高いと判定したDPCが元のDPCと、診断群によって決定される上位6桁まで一致した場合は「6桁的中」、入院目的や手術・処置の有無で決まる14桁全てで一致した場合

には「14桁的中」とした。

C. 研究結果

1) 索引語抽出

索引語の抽出解析では、約14000の疾患を特徴づける語句が抽出された。

表2に何種類のDPCで使用されていたかを語句別に示した。上位の語句は多くの疾患で共通的に使用されていることを示し、特に「入院」「退院」の2つは全てのDPCで使用されていた。

これらを出現頻度順に整列させると(図4)多くの疾患で共通に使用されている語句は少数であることが分かる。

2) DPC判定

DPCの判定では、「6桁的中」では24%、「14桁的中」では56%、合わせれば80%と高率にDPCを特定できた(図5, 表3)。的中できなかった症例は大きく2つに分けられた。(表4)1つは誤って判定したDPCが広い範囲に分散する場合で、「1602203x99x0xx」; その他の異常所見」などの様に定義が元々曖昧なDPC分類が多かった。もう1つは特定のDPCへ間違う場合で、「0201103x01x000」; 白内障、水晶体の疾患; 白内障手術+眼内レンズ挿入術」と「0202103x02xxxx」; 網膜血管閉塞症; 硝子体茎頭微鏡下離断術網膜付着組織を含むもの」の様に共通の術式を持つためキーワードが似通ってしまった場合と、「0501303x99000x」; 心不全」の様に原疾患と誤認する場合に更に分けられた。

D. 考察

我々はベクトル空間モデルを用いることで電子カルテから診療経過に関連する言語情報の抽出を行い、疾患を特徴付ける索引語の抽出や退院サマリーからの疾患特定の可能性を示してきた。今回の研究で、DPCにおいても、各分類を特徴付ける語句を抽出すること、逆にある語句がどのような分野で特異的にあるいは共通的に用いられているかを示すことが可能であった。その結果高い確率でDPCを特定し得たことは、本研究手法の有効性を改めて示すものであり、テキストマイニングは今後あらゆる医療文書に応用されていくと思われる。

今回得られた文書ベクトルは当院に固有のものであり、直ちに他院のサマリーから疾患を特定することは難しいと考えられる。しかし予備的な検討では同等の成績を示しており、今後更に対象疾患を広げると共に多数施設の症例を検討することで、普遍的なベクトル情報が得られる可能性を示したと考えられる。

今後の展開として、DPC判定の支援と監査を行うツールの作成を予定している。

E. 結論

テキストマイニング技術により、退院サマリーからDPCの決定と監査を支援できる可能性が示された。将来的には本手法を用いてDPC分類の見直しなどにも応用が可能であり、今後のDPCの精緻化や対象病院の拡大につれて本手法は重要性を増していくと考えられる。

F. 健康危険情報

本研究において健康危険事項は発生しなかった。

G. 研究発表

1. 論文発表

2. 学会発表

鈴木隆弘, 藤田伸輔, 石塚琳, 高林克日己. テキストマイニングによるDPC自動判定の試み. 医療情報学連合大会論文集 P686-687(2005)

(発表誌名巻号・頁・発行年等も記入)

H. 知的財産権の出願・登録状況(予定を含む。)

1. 特許取得

ありません

2. 実用新案登録

ありません

3. その他

ありません

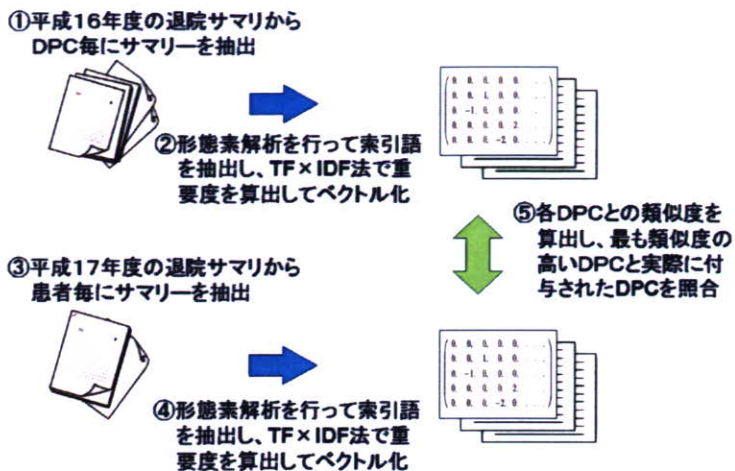
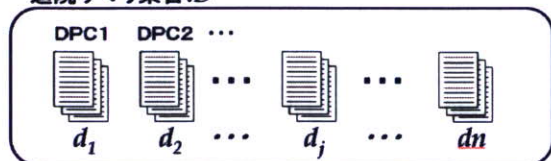


図1. 解析の手順

表1. DPC一覧 (抜粋)

DPCコード	DPC名称	症例数
0201103x01x000	白内障、水晶体の疾患	443
0400403x99x20x	肺の悪性腫瘍	402
0900103x010000	乳房の悪性腫瘍	174
1200103x99x10x	卵巣・子宮附属器の悪性腫瘍	173
0601603102xxxx	鼠径ヘルニア (15歳未満)	172
0500501xxxxxxx	狭心症、慢性虚血性心疾患 (検査入院)	160
0400403x99x00x	肺の悪性腫瘍	151
0601003x03xx0x	小腸大腸の良性疾患 (良性腫瘍を含む)	150
0705603x99x0xx	全身性臓器障害を伴う自己免疫性疾患	123
0400403x04x00x	肺の悪性腫瘍	116
1100801xxxxxxx	前立腺の悪性腫瘍 (検査入院)	100

退院サマリー集合: D



D より抽出した m 個の索引語 $w_1, w_2, \dots, w_i, \dots, w_m$

$$D = [d_1 d_2 \dots d_j \dots d_n] = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1j} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2j} & \dots & \alpha_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \alpha_{i1} & \alpha_{i2} & \dots & \alpha_{ij} & \dots & \alpha_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mj} & \dots & \alpha_{mn} \end{bmatrix}$$

あるDPC d_j における索引語 w_i の重要度

図2. ベクトル空間モデル

Term Frequency / Inverse Document Frequency

索引語の重要度

$$\alpha_{ij} = \frac{l_{ij} g_j}{n_i} \sqrt{\sum_{j=1}^m (l_{ij} g_j)^2}$$

局所的重み $l_{ij} = \log(1 + f_{ij})$

大局的重み $g_j = \log\left(\frac{n}{n_j}\right)$

文章正規化係数

- f_{ij} ... d_j における索引語 w_i の出現頻度
- n ... DPC総数
- n_i ... 索引語 w_i を含むDPC数

図3. Tf×Idf法による重要度算出

表2 語句別出現DPC数(上位30個)

順位	語句	出現DPC数
1	入院	126
2	退院	126
3	施行	124
4	受診	123
5	年	123
6	診断	123
7	外来	118
8	当科	118
9	紹介	117
10	後	115
11	病院	113
12	目的	112
13	開始	111
14	予定	109
15	当院	109
16	中	108
17	経過	108
18	左	107
19	所見	107
20	出現	106
21	上	104
22	経過観察	104
23	右	103
24	手術	103
25	治療	103
26	術後	103
27	検査	102
28	回	97
29	内服	95
30	入院時	94

表 4. 判定結果抜粋（上位）

DPCコード	DPC名称	同一DPC	同一診断群	的中せず	小計
0702203x021xxx	股関節症（変形性を含む。）；人工関節置換術 肩、股、膝（再置換でない場合）	28			28
0202103x97xxxx	網膜血管閉塞症；その他の手術あり	23			23
1101203x0401x	上部尿路結石症；体外衝撃波腎・尿管結石破碎術（一連につき）	23			23
1001803x99x00x	副腎皮質機能亢進症、非機能性副腎皮質腫瘍；なし	18			18
1100103x0101xx	腎の悪性腫瘍；腎（尿管）悪性腫瘍手術	16			16
0101553x99xx0x	脊髄性筋萎縮症および関連症候群；なし	13			13
0402603x99x0xx	原発性肺高血圧；なし	12			12
1000203x99x2xx	甲状腺の悪性腫瘍；なし	11			11
1200203x99x21x	子宮頸・体部の悪性腫瘍；なし	9			9
0600203x99x11x	胃の悪性腫瘍；なし	8			8
総計		1800	771	616	3187

表 5. 判定結果抜粋（下位）

DPCコード	DPC名称	同一DPC	同一診断群	的中せず	小計
0603453x99x00x	胆管炎；なし	4		8	12
1400303x9900xx	周産期に発生した新生児の障害；なし	11		23	34
1400103299000x	妊娠期間短縮、低出産体重に関連する障害（出生時体重1500g以上2500g未満）；なし	7		15	22
0601303x99x00x	食道、胃、十二指腸、他腸の炎症（その他良性疾患）；なし	6		14	20
1610603x97x0xx	詳細不明の損傷等；あり	1	1	6	8
0800113x99xxxx	急性膿皮症；なし	1		5	6
0700403x97xxxx	骨軟部の悪性腫瘍（脊椎脊髄を除く。）；その他の手術あり	1		6	7
1601603x99x00x	敗血症その他の感染症；なし	1		15	16
1103203099xxxx	腎、泌尿器の疾患（その他）（6歳以上）；なし			9	9
1602203x99x0xx	その他の異常所見；なし			67	67
総計		1800	771	616	3187