

厚生労働科学研究費補助金

第3次対がん総合戦略研究事業

がん予防に有用な情報基盤整備に関する研究

(H16-3次がん-011)

平成16年度 総括・分担研究報告書

主任研究者 小山 博史

平成17(2005)年3月

# 目 次

I. 総括研究報告	
がん予防情報処理基盤に関する調査研究	----- 1
小山博史	
II. 分担研究報告	
1. がん予防に関する知識の体系化に関する調査研究に関する研究-----	14
小野木雄三	
2. 臨床疫学手法を用いたがん予防情報解析アルゴリズム開発に関する研究に関する研究-----	17
小出大介	
3. がん予防対象患者の病院情報システムからの抽出法に関する調査研究に関する研究-----	20
大江和彦	
4. がん検診画像データからの予防情報抽出に関する研究に関する研究-----	23
若尾文彦	
5. がん予防薬の臨床試験支援用情報システム構築に必要となる機能仕様に関する調査研究に関する研究	----- 26
山本精一郎	
6. がん予防薬の薬物動態関連酵素とSNPとの関連データベースの開発-----	33
日紫喜光良	
III. 研究成果の刊行に関する一覧表	----- 38
IV. 研究成果の刊行物・別刷	----- 46

## I. 総括研究報告書

厚生労働科学研究費補助金(第3次がん総合戦略研究事業(分野3))

(主任)研究報告書

研究課題名:「がん予防情報処理基盤に関する調査研究」

主任研究者氏名:小山 博史

所属:東京大学大学院医学系研究科

研究要旨: 本年度は、がん予防情報処理基盤として①がん予防の情報基盤の一つである日本語対応のがん予防情報データベースの論理設計の基本となるデータベース項目について検討をおこない、NCIのオントロジを基にがんに関連する食品と化学物質のオントロジの日本語化を行った。②がんの三次予防への応用目的でがん予防薬の遺伝的多型情報を加味した薬物動態シミュレーションモデルの最適化手法に関する検討を行った。③がんの一次予防を目的として一般向けがん予防情報提供に関する利用者の検索特性に応じたがん予防情報の最適検索システム“CaCheBroker”の試作を行った。④がん予防の個別化を実現を目的として、一般利用者の特性に応じたがん予防食品などの情報提供や化学がん予防薬の臨床試験を進めるための遺伝的多型も加味したがん予防薬の最適化手法へのグリッド技術の応用についての検討を進めた。

東京大学大学院医学系研究科  
クリニカル・バイオインフォマティクス研究ユニット, 特任教授  
小山博史  
東京大学大学院医学系研究科  
医療情報経済学講座, 教授  
大江和彦  
東京大学大学院医学系研究科  
クリニカル・バイオインフォマティクス研究ユニット, 特任教授  
児玉安司  
東京大学大学院医学系研究科  
クリニカル・バイオインフォマティクス研究ユニット, 特任助教授  
小野木雄三  
国立がんセンター中央病院  
放射線診断部, 医長  
若尾文彦  
東京大学大学院医学系研究科  
クリニカル・バイオインフォマティクス研究ユニット, 特任助教授  
小出大介  
国立がんセンター研究所  
がん予防・検診研究センター情報研究部予防  
・検診情報評価室, 室長  
山本精一郎  
産業技術総合研究所  
生物情報解析研究センター, 主任研究員  
日紫喜光良

## A. 研究目的

本年度は、がん予防の個別化を実現を目的として、一般のがん予防情報検索者の特性に応じたがん予防食品などの情報提供、化学がん予防薬の臨床試験を進めるためのがん予防薬の遺伝的多型も加味した最適化手法についての検討を進めた。特に、がん予防情報処理基盤としてがん予防の情報基盤の一つである日本語対応のがん予防情報データベースの論理設計の基本となるデータベース項目としてNCIのオントロジを基にしたがんに関連する食品と化学物質の日本語化を行った。また、がんの三次予防への応用の目的で現在開発中のがん予防薬の遺伝的多型情報を加味した薬物動態シミュレーションシステムを用いたがん予防薬剤の選択に必要な情報処理に関する検討、がんの一次予防を目的とした一般向けがん予

防情報の個別最適化手法を開発することを目的とした。

## B. 研究方法

①NCIのオントロジを基にしたがんに関連する「化学薬品・薬剤」概念の日本語化。

②がん予防薬の個別化最適化を支援するための薬物動態モデルの開発。

③一般向けがん予防情報の検索者の特性を加味したSearch Engine Optimizationに関するがん予防情報に特化した検索システム「がん予防情報提供システム」(CaCheBroker)の開発。

上記の方法と結果については以下の研究結果にまとめて述べる。

(倫理面への配慮) 本年度の研究では、特に個人に関わる情報の処理解析対象データは取り扱わなかった。

## C. 研究結果

(1)NCIのオントロジを基にしたがんに関連する「化学薬品・薬剤」概念の日本語化:

がん予防情報の基盤の中でもがん予防に利用されている日本語の用語体系化は、情報検索やデータベースの構築において重要である。その理由の一つは、がんの一次予防法としては喫煙や飲酒などの生活習慣情報や食事療法、サプリメント情報など既存の医学用語の範囲とはことなる用語が使用されていることが予想される。この問題に関する検討は今後

の課題であるが、本年度は大阪大学溝口研で開発されたオントロジエディタ「法造」ver4.0を用いてNCIのオントロジの中でもがん予防と関連の深い「化学薬品・薬剤」の日本語化を行った。その結果として大項目としては、「化学薬品・薬剤の機能」と「化学薬品・薬剤の機能構造」、「使用しなくなった化学薬品・薬剤の機能」の3つで構成した。「化学薬品・薬剤機能」は、「薬理物質」「化学修飾」「臨床試験要素」「生理学的調整因子」「色素沈着」「食物・食品」「薬物乱用」「試薬」「産業生成物」「競争阻害薬」「構造類似体」に分類されていた。また、「薬剤・化学薬品構造」は、大きく「有機化合物」「非有機化合物」に分かれており、「有機化合物」は、「アミノ酸蛋白質」「芳香環化合物」「脂質」「有機金属・隣・ホウ素誘導体」「炭化水素」「炭水化物」「核酸・ヌクレオチド・ヌクレオシド」「マレイン酸」「メシル酸」「有機オキソ酸」「アルカロイド」「アルコール化合物」「ラクトン化合物」「メチルエーテル」「有機塩素化合物」で構成されていた。非有機化合物質としては、「元素・同位元素」「フリーラジカル」「イオン」「塩」「鉱物」「電解質」「臭化シアン」で構成されていた。例えば「クルクリン」や「フラボノイド」は、「バイオフィラボノイド」とis-a関係にあり「バイオフィラボノイド」は「フラボノイド」とis-a関係にあり、「フラボノイド」は「栄養製剤」とis-a関係にあり、「栄養製剤」は「食品構成要素」とi

s-a関係にある。この関係を基に検索やデータベースの論理設計への利用を目指した。

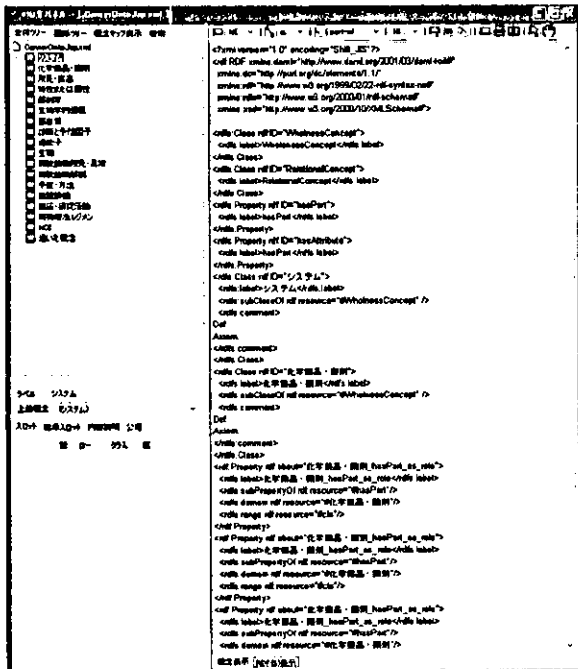


図1. オントロジーエディタ「製造」を用いたがん化学薬品・薬剤オントロジの日本語化.

(2) がん予防薬の個別化最適化を支援するための薬物動態モデルの開発:

CellMLを基本とした遺伝的多型を基に薬物の代謝反応速度係数を変化させて臓器コンパートメント毎の動態をシミュレーションできる作成ツールを試作し、薬物動態解析モデルのパラメータフッティングについて検討した。薬物動態モデルの作成には、特に薬物投与のタイムコースの自由度を高めた。任意の時刻での急速または一定速度での薬物投与を設定可能にした。薬物動態モデルに、被験者の個体

情報の依存性を加味する機能を付け加えた。加味する被験者の個体情報とはBSA、薬物感受性遺伝子関連酵素の発現に関与する遺伝子多型情報(SNP)とした。被験者の個体情報は、薬物動態モデルを各被験者に対してパラメータフッティングを行い、フッティングしたパラメータの被験者の個体情報への依存性をもとに定めた。作成した薬物動態モデルの評価を行った。薬物動態のモデリングをより素早く行うために、既存の薬物動態を記述するエディタに機能として階層構造の表示・変数の対応づけの詳細な表示を可能とした。測定によって得られたPKデータと数式上の変数を対応づける機能を追加した。対応づけはCSVファイルによって行うこととした。測定によって得られた個体情報と数式上の変数を対応づける機能を追加し、CSVファイルによって行うこととした。

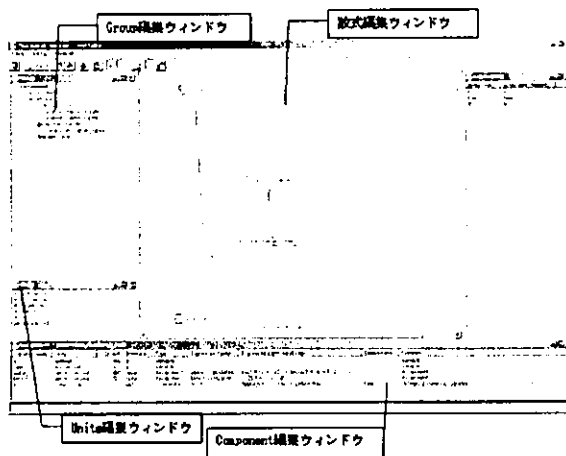


図2. CellMLを基本に遺伝的多型を基に薬物の代謝反

応速度係数を変化させて臓器コンパートメント毎の動態をシミュレーションできる作成ツール。

設定されたPKデータ、個体情報データ、数式を統計ソフトウェアRの形式で出力する機能を追加し、そのデータをもとにRを用いてパラメータフィッティングを行った。

アプリケーションは、①薬物動態を記述する数式を作成するためのアプリケーションと②仮想的な患者の個体データと薬物投与でシミュレーションした。薬物動態を記述する数式を作成するためのアプリケーションは、コンパートメントモデルの作成・数式の個体情報への依存性を記述した。

コンパートメントモデルの段階の数式には被験者の個体情報を入れなかった。例えば、1コンパートメントモデルの場合、

$$\frac{dC}{dt}(t) = \frac{\text{Dose}}{Vd} * \exp[-t * \frac{Cl}{Vd}]$$

という数式が相当する。分布容積VdやクリアランスClなどの被験者の個体情報への依存性は、PKデータと照らし合わせて決定した。アプリケーション機能は、①グラフによる数式の記述する機能、②各ノードはマウス操作により自由に追加／削除する機能、③各ノードに変数(物質質量)、反応速度定数、微分方程式を記述する機能、④数式のCellML形式での保存する機能とした。パラメータフィッティングを行う

ために、測定されたPKデータ内の項目と数式上の変数を対応づけた。PKデータ内の薬物の血中濃度と、数式の血中濃度を表す変数との対応づけを行った。アプリケーションは、利用者が定義したPKデータ、PKデータの項目と数式上の変数との対応関係、数式のRスクリプトにより出力とした。これらのRスクリプトを用いて臨床データを使ってパラメータフィッティングを行うことを可能とした。開発したアプリケーション機能は、①指定したPKデータの読み込む機能、②指定したCellMLの読み込む機能、③PKデータの項目と数式上の変数との対応づける機能、④PKデータと数式および項目—変数対応づけのR対応形式での出力機能である。

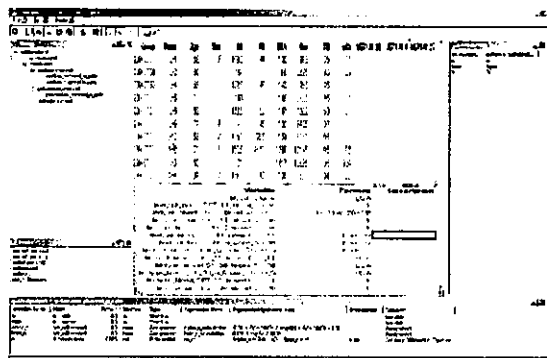


図3. 薬物の代謝反応速度係数の最適化処理用画面例。

数式上の変数へ被験者の個体情報への依存性を記述する。例えば、1コンパートメントモデルの場合、

$$\frac{dC}{dt}(t) = \frac{\text{Dose}}{Vd} * \exp[-t * \frac{Cl}{Vd}]$$

とした。

R上のパラメータフィッティングにより、数式上の変数CIは被験者の体重(Wt)とクレアチニン・クリアランスCcrの一次関数になっていることがわかったとすると

$$CI = \exp[a * Weight + b * Ccr + c]$$

となる。

ここで、定数a,b,cはR上でのパラメータフィッティングにより得られた定数である。この記述をシミュレーションに反映するためにユーザの指定したCellMLの読み込む機能、被験者の個体情報への依存性を追加した数式の作成する機能、被験者の個体情報への依存性を追加した数式のCellML形式での保存する機能を開発した、

遺伝的多型のパターン毎にシミュレーションできる機能として、指定された変数の値変化をファイル(CSV形式)に保存する機能をエディタから出力するRスクリプトにユーザに指定された変数をCSV形式で出力コマンドにより実現した。また、ユーザに指定された変数の値変化をプロットする機能として、エディタから出力するRスクリプトに指定された変数をプロットするためのコマンドを開発した。さらにシミュレーションを行う時間を設定する機能、値変化を保存する変数名を設定する機能、値変化をプロットする変数名を設定する機能、値変化を保存するCSV形式のファイル名を設定する機能

を開発した。次年度は、これを用いてSNPの違いによる個別のがん予防薬の薬物動態シミュレーションモデルの作成とそれを用いたパイオマーカの探索を行うことを計画している。

(研究協力者:安藤雄一(埼玉医科大学)、白瀉宏之、倉田正)

(3) 一般向けがん予防情報の検索者の特性を加味したSearch Engine Optimaizationに関するがん予防情報検索システム「がん予防情報提供システム」(CaCheBroker)の開発。

本システムは、WWWインターフェースを介して、利用者のがん予防情報を提供するWebサーバアプリケーションにより構成した。利用者は、「利用者情報管理システム」によって、パブリック認証、または、個人認証の元のがん予防情報提供システムにアクセスすることとした。利用者からの要求を受けたがん予防情報提供システムは、「検索語分類システム」、「検索語変換システム」によって要求を解析し、がん予防情報関連データベースや、インターネットから情報を取得し、「コンテンツ生成システム」によってコンテンツを生成し利用者に提供することとした。利用者によるシステムの利用状況は、「統計情報蓄積システム」により蓄積される。情報管理者は、管理者認証の元のがん情報提供システムにアクセスし、「統計情報システム」によって、がん予防情報関連データ



ベースに蓄積された利用者情報を取得可能とした。

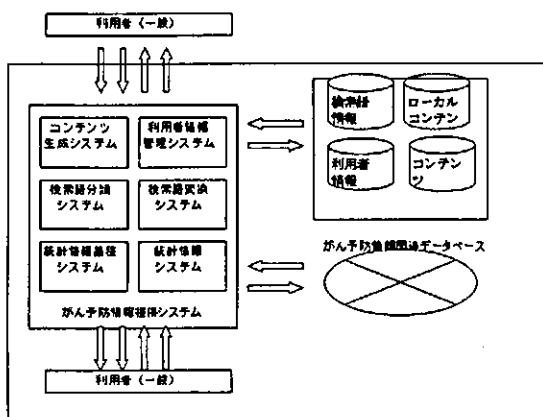


図3. 個別最適化がん予防情報提供システム構成図。

「がん予防情報提供システム」の稼働環境を下記に示す。

動作オペレーションシステム:

- OS: UNIX (Linux)
- データベースマネジメントシステム  
PostgreSQL v7
- ソフトウェア開発言語  
Java, Perl5 など
- 動作システム環境  
主記憶: 512MB以下  
ハードディスク: 100GB以下(ただし、システム動作に必要な容量)

「がん予防情報提供システム」の目的は、個々の利用者の特性データの収集とそれに応じた最適ながん予防情報の提供を実現することにある。そのために以下のデータを処理対象とし

た。システム利用者プロフィールデータ:利用者1,000名以上、管理者10名以上による利用が可能とした。ただし、将来の運用に際し、利用者1,000,000名以上への拡張が可能とした。利用者プロフィールデータは、個人情報保護法に抵触しないように個人を特定情報以外の個別特性を登録可能な個人特性データとシステム利用データ別に解析可能なデータ構造とした。検索語情報登録データは、50,000以上の用語(日本語)を処理可能とした。ローカルコンテンツとして、東京大学大学院医学系研究科臨床バイオインフォマティクス研究ユニット・臨床情報工学部門で作成されたデータより10,000ページ以上に相当するコンテンツデータを対象とし、その登録・検索・参照・管理ができることとした。本年度は、3冊のがん予防に関する書物の要約とキーワードを有するスライド約300枚を作成した。

ポータルサイトからの論文データベース検索としてPubMedの利用を可能とし、Webページの検索エンジンとしてGoogleの利用を可能とした。利用者の要求情報を蓄積し、個別又はセッション単位の利用と、提供したコンテンツの情報を蓄積し、利用者の要求情報と共に利用可能とした。利用者管理システムは、利用者および管理者の認証、利用者の特性情報の登録と管理、利用者および管理者のシステム利用履歴の管理を可能とした。特性情報

としては、身体情報(年齢、身長、体重など)、疾病・治療履歴等個別最適化に必要となる情報項目を登録できるようにした。また、それ以外の目的でも必要に応じて登録できる内容を追加しやすいシステム構成とした。本システム運用においては、個人特性情報は個人情報保護法に抵触しないように個人を特定できる情報は取り扱わず、個別特性解析は非連結匿名化し独立したサーバにて管理し、セキュリティーの確保を可能とした。

検索語分類システムは、今回作成したがん予防に関する要約コンテンツより用語(予約語)を抽出し、予約語間の共起情報および出現頻度情報と共にデータベース化した。検索語変換システムは、利用者の入力した検索語を逐次解析し、利用者が求める情報の検索に適切と思われる予約語集合のリストアップを可能とした。利用者の入力した検索語の予約語への対応は、検索語分類システムによって構築された用語データベースの検索により行える。用語データベースに存在しない検索語に関しては、検索エンジン(PubMed又はGoogle)を解して関連する予約語を探索可能とした。統計情報蓄積システムは、利用者が検索開始から目的のコンテンツが見つかるまでに、検索に用いた予約語とホームページのURLとその順番を検索時間と共に保存できることとした。蓄積した統計情報により、利用者が入力した

検索語あるいは選択した予約語から最終的にたどり着くと予想されるコンテンツをリストアップできることとした。検索ログ・検索語解析システムは、利用者が入力した検索語に対応した検索用予約語集合を検索語分類システムによって構築した用語データベースより抽出し表示することとした。表示した検索用予約語集合から検索を行う用語を選択することにより、検索を実行し、検索実行後、結果一覧を表示するとともに、蓄積した統計情報を利用して検索に用いた用語から最終的にたどり着くと予想されるコンテンツの一覧の表示も可能とした。検索実行結果一覧または予想されるコンテンツの一覧で表示されたリンクを選択することでコンテンツ本文を表示できることとした。表示したコンテンツ本文が目的のものであるか否かの評価を利用者がワンクリックで行うことを可能である。統計情報を作成するシステムは、システムの利用状況及びシステムに蓄積された利用者情報や用語情報を管理者が閲覧する機能を開発した。システム利用状況は、システムを利用中のユーザのリストと動作、指定期間における利用状況を表示可能とすることとし、指定した利用者の特性情報と利用履歴を表示可能とすることとした。用語情報としては、利用者が目的のコンテンツにたどり着くまでに指定した予約語とその順番を統計情報として表示可能とすることとした。

次年度は、このシステムを試験的に運用し、利用者の特性抽出し、その特性に合った情報提供の最適化法の開発を目指す予定にしている。

(研究協力者:伊藤勝彦、日紫喜光良(産業技術総合研究所))

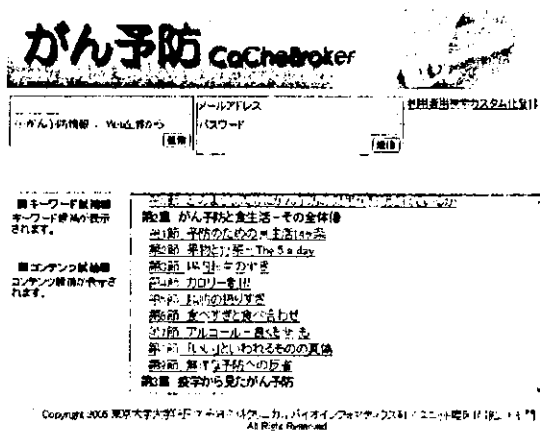


図4. がん予防情報個別最適化システムのプロトタイプシステムの外部設計画面(案).

#### D. 考察

現在、がん予防に関する情報がインターネット上に激増している。また、利用者があるWebサイトへ来訪するには、今までメールや広告、各種メディア、Webページなどからのリンクや紹介が多かったが、最近では検索エンジンの結果をたどってくるケースも非常に増えている。そのため、検索エンジンの結果で上位に表示されることは、WebサイトやWebページの閲覧数を増やし、情報へのアクセサ

ビリティを決定させる重要な要素となっている。また、膨大な検索結果から利用者が求めるコンテンツを得ることへの困難性は日々高まっており、

本研究では、がん一次予防を目的とした「一般向けがん予防情報の最適検索システム“CaCheBroker”」を試作したが、これについてはがん予防情報のポータルサイトとしてのホームページを立ち上げを早期に実現する計画である。また、他のSearch Engineとの比較試験を予定している。さらに、がん予防に関するオントロジーを適応し日本語でのMEDLINE文献やインターネット上の情報を検索するシステムへと開発を進める。これが実現すれば、今まで日本語でヒットしなかったページへのアクセスも可能となり、海外のがん予防に関する最新情報を収集することが可能となる。

今回は、一次予防情報の中で主にテキスト情報の提供に関する提供に関する研究を進めてきた。今後評価委員会からのコメントを再度検討し、今回作成した情報提供システムに動画も含めた臨床的な有用性に関する検証を行うことを予定している。

がんの3次予防支援を目的とした遺伝的多型情報に応じたがん予防薬投与量推定に関する個別化に関する研究を進めている。今後大腸がんとアスピリンなどのがん予防薬の薬物動態解析用のモデル化について検討を進める

予定にしている。

## E. 結論

本年度は、①がん予防の情報基盤の一つである日本語対応のがん予防情報データベースの論理設計の基本となるデータベースに必要な項目について検討をおこない、NCIのオントロジを基にしたがんに関連する食品と化学物質の日本語化を行った。②がんの三次予防への応用の目的で開発中のがん予防薬の遺伝的多型情報を加味した薬物動態シミュレーションモデルの最適化手法に関する検討を行った。③がんの一次予防を目的とした一般向けがん予防情報の個別最適化手法の開発を目指したがん予防情報検索システムを作成した。がん予防情報提供に関する利用者の検索特性に応じたがん予防情報の最適検索システム“CaCh<sub>3</sub>Broker”の試作を行った。④がん予防の個別化を実現を目的として、一般利用者の特性に応じたがん予防食品などの情報提供や化学がん予防薬の臨床試験を進めるためのがん予防薬の遺伝的多型も加味した最適化手法へのグリッド技術の応用についての検討を進めた。

## F. 健康危機情報 なし

## G. 研究発表

### 1. 論文

1) Kuroda Y, Nakao M, Kuroda T,

Oyama H and Komori M, “Interaction Model between Elastic Objects for Haptic Feedback considering Collisions of Soft tissue”, Computer Methods and Programs in Biomedicine(Elsevier Science),2005 (To appear)

- 2) Kuroda Y, Nakao M, Kuroda T, Oyama H. and Yoshihara H., “Shape Perception with Friction Model for Indirect Touch”, World Haptics Conference (IEEE Proceedings), 2005 (To appear) Y. Kuroda, M. Nakao, T. Kuroda, H. Oyama and H. Yoshihara, “MVL: Medical VR Simulation Library”, Proc. Medicine Meets Virtual Reality 13 (MVR13), pp.273-276, 2005.
- 3) 小山博史: オーダーエントリーシステムと電子カルテ, Surgery Frontier Vol.12, No.1, 2005.
- 4) 中尾 恵, 黒田知宏, 小山博史, 湊小太郎, “術前VRリハーサル”, 第17回VR-Lab. シンポジウム, 東京大学, Mar 2005.
- 5) Nakao M, Imanishi K, Kuroda T, Oyama H. Practical haptic navigation with clickable 3D region input

- interface for supporting master-slave type robotic surgery. Stud Health Technol Inform. 2004; 98:265-71.
- 6) Nakao M, Kuroda T, Komori M, Oyama H, Komeda M. Physics-based preoperative approach planning using hybrid virtual bodies. Stud Health Technol Inform. 2004; 98: 262-4.
- 7) Kuroda Y, Nakao M, Kuroda T, Oyama H, Komori M, Matsuda T. FEM-based interaction model between elastic objects for indirect palpation simulator. Stud Health Technol Inform. 2004; 98: 183-9.
- 8) Kuroda Y, Nakao M, Kuroda T, Oyama H and Yoshihara H, "Haptic Rendering Method of Friction Forces for Indirect Shape Perception", Proc. 10th International Conference Virtual Systems of MultiMedia, pp. 232-238, 2004.
- 9) Matsuya S, Onogi Y, Shinohara N, Yamaguchi I, Watanabe H, Ohe K, Yamaguchi K, Niidome T, Oyama H. Physician order entry of ultrasound examination with handheld wireless terminal :Medical Imaging 2004 PACS and Imaging Informatics, 43-51, 2004.
- 10) Nakao M, Imanishi K, Kuroda T, Oyama H, Practical Haptic Navigation with Clickable 3D Region InputInterface for Supporting Master-Slave Type Robotic Surgery: Medicine Meets Virtual Reality 12 J.D.Westwood et al.(Eds.)IOS Press, 265-271, 2004
- 11) Kuroda Y, Nakao M, Kuroda T, Oyama H, Komori M, Matsuda T, FEM-based Interaction Model between Elastic Objects for Indirect Palpation SimulatorMedicine Meets Virtual Reality 12 J.D.Westwood et al. (Eds.) IOS Press, 183-189, 2004.
- 12) 黒田嘉宏, 中尾 恵, 黒田知宏, 小山博史, 吉原博幸, "高度な触診スキルの習得を目的とした摩擦力提示手法", 計測自動制御学会 システムインテグレーション部門 講演会論文集, pp. 370-371, 2004.
- 13) 黒田嘉宏, 中尾 恵, 黒田知宏, 小山博史, 松田哲也, 吉原 博

- 幸, "MVL: 実時間医用VRシミュレーションライブラリの開発", 日本バーチャルリアリティ学会 第9回大会, pp. 533-536, 2004.
- 14) 黒田 嘉宏, 中尾 恵, 黒田 知宏, 小山博史, 小森 優, 松田 哲也 吉原 博幸, "間接触診による形状触知VRシミュレーション", システム制御情報学会, 2004.
- 15) 松本伸哉, 今村知明, 神奈川芳行, 田島文一, 松谷司郎, 小山博史:医療における関連性分析を用いた三徴候の識別方法,第5回計測自動制御学会システムインテグレーション部門講演会,2004.12.17-19.
- 16) 陳俊成, 小山博史, 菅野一也, 大江和彦: 伝播モデルを用いる細菌検査結果のトランザクションデータの解析に関する研究,第5回計測自動制御学会システムインテグレーション部門講演会,2004. 12. 17-19.
- 17) 義澤宣明, 船曳淳, 小山博史: リスクコミュニケーションにおける医療情報システムの役割,第5回計測自動制御学会システムインテグレーション部門講演会,2004.12.17-19.
- 18) 篠原信夫, 小山博史, 松谷司郎, 大江和彦: 合併症発生疑い事例の医療情報データを用いた抽出法の開発,第5回計測自動制御学会システムインテグレーション部門講演会,2004.12.17-19.
- 19) 篠原信夫, 小山博史, 松谷司郎, 大江和彦: 病院情報システムデータのみからの医療事故発生疑い患者抽出手法,第24回医療情報学連合大会(第5回日本医療情報学会学術大会),2004.11.26-28
- 20) 松谷司郎, 小山博史, 篠原信夫: 医療情報処理におけるSQL-92とATSQL2の「時制」の扱いやすさの比較,第24回医療情報学連合大会(第5回日本医療情報学会学術大会),2004.11.26-28.
- 21) 黒田嘉宏, 中尾恵, 黒田知宏, 小山博史, 松田哲也, 吉原博幸: 計状触知を可能とする弾性体を対象とした摩擦の力学レンダリング手法,日本バーチャルリアリティ学会第9回大会論文抄録集,82,2004.
- 22) 原田雅之, 高橋修一, 寺田尚文, 黒田知宏, 小山博史: 教育用遠隔共用VR型手術シミュレーター構築,日本バーチャルリアリティ学会第9回大会論文抄録集, 100, 200

4. 日本職業・災害医学会会誌, 第52卷, 第2号, 91-95, 2004.
- 23) 黒田嘉宏, 中尾恵, 黒田知宏, 小山博史, 松田哲也, 吉原博幸: MVL:実時間医用VRシミュレーションライブラリの開発, 日本バーチャルリアリティ学会第9回大会論文抄録集, 100, 2004.
- 24) 小山博史: 医療のIT化の未来, MEDICAL DIGEST Vol.53, 通巻392, 53-60, 2004.
- 25) 小山博史: 臨床情報工学と生体工学, Surgery Frontier Vol.11, No.3, 2004.
- 26) 小山博史: 電子カルテとは?,
- 27) 小山博史: 病院情報システム更新の考え方と院内体制, ITVISION No.5: 17-19, 2004.
- 28) 小山博史: ITと医療: 情報から知識へ, 教育と医学 No.607: 41-48, 2004.

## 2. 学会発表 準備中

## H. 知的財産権の出願・登録状況 (予定を含む)

1. 特許取得 予定
2. 実用新案登録 なし
3. その他 なし

## II. 分担研究報告書



厚生労働科学研究費補助金(第3次がん総合戦略研究事業(分野3))

(分担)研究報告書

研究課題名:「がん予防に関する知識の体系化に関する調査研究」

分担研究者氏名: 小野木 雄三

所属: 東京大学大学院医学系研究科・臨床バイオインフォマティクス研究ユニット・臨床情報工学

研究要旨: 研究者や一般大衆を対象としたがん予防に関する知識を収集するために、検索のためのオントロジー構築を目指し、まず既存のオントロジーであるSNOMED-CTとUMLSのSemantic Networkを利用することを検討した。結果としてSNOMED-CTのサブセットである物質と、それに効果の下位概念として誘因・抑止を加えることによりオントロジーを構築できることが明らかになった。またSemantic Networkを利用することにより、半自動的に物質とがんと効果を記述できることが示された。またSNOMED-CTの概念数30万のうち、日本語で記述可能な概念数はわずかに12000件程度であった。

#### A. 研究目的

研究者や一般大衆を対象としたがん予防に関する知識を収集するために、検索のためのオントロジー構築を目的とした。

#### B. 研究方法

米国NLMにより公開されているUMLSを入手し、それに含まれるSemantic Network、SNOMED-CTを利用した。なおSNOMED-CTを利用するには別に米国病理学会からライセンスを取得した。米国Stanford大学で開発されたProtege を利用してオントロジー開発を行った。また日本語との対応はUMLS中のMeSH日本語版を利用し、UMLS全体を収容したデータベース上の操作により行った。

#### C. 研究結果

がん予防には、将来がんになるリスクを有する物質への暴露を少なくする、早期発見に努める、積極的にがんを予防する物質を摂取する、さらに進んでがん予防薬を使用する(がん再発を防止するものも広義の予防薬と考えられる)があげられる。これらに関する遺伝子解析・蛋白質解析などの技術により、がん予防に関連する知識は従来にない勢いで増大しつつある。また学術的に RCT を経た EBM によって確立された知識だけでなく、サプリメントや代替療法などの知識まで含めるならば、さらに膨大な知識量となっている。これらの膨大な知識に新たなひとつを加えるアプローチも重要であるが、既存の知識を整理し何が行われているのかを効率的に検索することも極めて重

要である。新たな知見をひとつ加えるにも、まずは既存知識の中に何があるのかを知り、すでにどこかで明らかになっている知見ではないことを確認しなくてはならないからである。こうした知識の利用者は臨床家、研究者、そして一般大衆が考えられる。臨床家は EBM だけを利用すると考えられ、研究者は EBM に限らず広く文献情報を利用する。いずれも英語での知識で構わないし検索も英語で十分である。しかし一般大衆は日本語での利用が主体となる点を考慮すべきである。

がん予防に関する知識は、MEDLINE などの文献から得られるほか、インターネット上に存在するがん予防に関する情報を収集している権威あるサイトなどから得ることができる。また代替療法に関してもサイトが存在するほか、インターネット上の膨大なテキストから知識を収集することができる。いずれにしても、これらの知識を収集するには「何らかの物質ががんに対して何らかの効果を持つこと」を調べることになる。この際に物質とは何か、および効果とは何か、そしてがんとは何か、を明確に記述できていることが極めて重要である。これらの概念を、階層構造と相互関係性の総体として定義した上で、個々の概念を記述する表記を列挙したオントロジーを構築することが重要であり、このオントロジーなくしてがん予防に関する知識を効率的に検索することは難しい。

昨年度は文献情報を人間が個々に確認して CAT (Critically Appraised Topic) を作成し情報を蓄積した。これは臨床家を対象としたものである。しかし今年度は研究者や一般大衆を対象としたがん予防に関する知識を収集するために、検索のためのオントロジー構築を目指し、まず既存のオントロジーである SNOMED-CT と UMLS の Semantic Network を利用することを検討した。SNOMED-CT では、がん予防に関連する物質は is\_a 関係で記述されている。SNOMED-CT には is\_a 以外に多様な関係が導入されているが、がんとこれに関連する物質との間に特有の関係性はほとんど記述されていない。従って SNOMED-CT のサブセットである物質と、それに効果の下位概念として誘因・抑止を加えることによりオントロジーを構築できることが明らかになった。また UMLS の Semantic Network には、物質とがんとの関係性が記述されているので、手動的にではなく半自動的に物質とがんとの効果を記述できることが示された。また日本語との対応に関し、すでに MeSH に対応する日本語医学用語 5 万語が存在しているが、SNOMED-CT の概念数 30 万のうち、日本語で記述可能な概念数はわずかに 12000 件程度であることがわかった。日本語に関しては、今後医学用語に関する英和辞典などを活用して対応関係を増やす必要がある。

## D. 考察

本研究の成果は「がん予防情報検索のためのオントロジー」構築であるが、これを利用して実際のMEDLINE文献やインターネット上の情報を検索するシステムが開発可能である。これが実現すれば、オントロジーを随時更新することにより、新しい文献やインターネット上の情報を検索して、がん予防に関する最新情報を収集することが可能となる。収集した情報は、インターネットを通じて公開する予定である。

## E. 結論

研究者や一般大衆を対象としたがん予防に関する知識を収集するために、検索のためのオントロジー構築を目指し、まず既存のオントロジーであるSNOMED-CTとUMLSのSemantic Networkを利用することを検討した。結果としてSNOMED-CTのサブセットである物質と、それに効果の下位概念として誘因・抑止を加えることによりオントロジーを構築できることが明らかになった。またSemantic Networkを利用することにより、半自動的に物質とがんと効果とを記述できることが示された。またSNOME

D-CTの概念数30万のうち、日本語で記述可能な概念数はわずかに12000件程度であった。

## F. 研究発表

### 1. 論文

- 1) 小野木雄三, UMLS を利用した日本語医学知識サポートシステム. 医療情報学, 24(suppl.), pp386-387, 2004.
- 2) Onogi Y, Ohe K, Tanaka M, Nozoe A, Sasaki T, Sato M, Kikuchi Y, Shinohara T, Suzuki H, Kaihara S, Seyama Y: Mapping Japanese Medical Terms to UMLS Metathesaurus. in Proceedings of the IIth World Congress on Medical Informatics. Sep.7-11,2004, San Francisco, California, USA. (EDS) M. Fieschi et al. p406-410, 2004.

### 2. 学会発表 準備中

## G. 知的財産権の出願・登録状況 (予定を含む)

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

厚生労働科学研究費補助金(第3次がん総合戦略研究事業(分野3))

(分担)研究報告書

研究課題名:「臨床疫学手法を用いたがん予防情報解析アルゴリズム開発に関する研究」

分担研究者氏名: 小出 大介

所属:東京大学大学院医学系研究科・臨床ハイパフォーマンス研究ユニット・臨床疫学部門

研究要旨:がん予防薬の薬理遺伝学情報データベースの基礎となる予防薬を検索してリスト化した。リストには薬物名, 代謝酵素名, 遺伝子多型情報, 有害事象, 病名等を含め、その中で使用される特に病名(効能・効果や有害事象名など)について国際標準のMedDRAに基づいてコード化し、薬物名についてはまだ国際標準がないことから国内標準のHOT7を用いた。がん予防情報の解析についてはEBMの手法に基づいた質的評価も加えた。

## A. 研究目的

がん患者の増加とがん予防への具体的な施策が社会技術として求められていることから、がん予防情報やがん予防薬の効果予測法の開発に関する研究と予防情報のデータマイニングやグリッド技術を用い効率的に行う情報処理技術等を開発することが全体の目的であり、その中で本分担研究は、EBM等で注目を集めている臨床疫学の手法を用い、がん予防情報を解析するためのアルゴリズム開発をすることである。16年度ではインフラストラクチャーとなるデータソースの構築を行い、今後の解析を視野に入れて標準的なコード化を行った。なおがん予防薬については1次～3次予防の全てを含む広範な範囲で捉えた。そしてサブリメントなどの情報も取り入れることとした。

## B. 研究方法

情報源としては、PubMedで代表される米国立医学図書館(NLM)の文献データベース、The Cochrane Libraryのデータベース、日本の文献としては、医学中央雑誌のデータベース、がん治療に用いられる医薬品添付文書を利用した。

病名については、国際標準の薬事統制用語集(MedDRA)によってコード付けを行い、医薬品名については国際標準が存在しないことから、国内標準であるHOT7によるコード付けを行った。収集した情報は、EBMの手法に則った質的評価を行い、最も質が高いとされる「無作為化比較試験の系統的レビュー」のレベル1aから、最も低い「一連の症例研究」のレベル4までのエビデンスレベルをふった。記述表現はXMLを用いた。構築されるデータソースの項目としては、薬物名、代謝酵素名、遺伝子