

制約自然言語による医療知識基盤の構築

Development of the Medical Knowledge Infrastructure by the Formal Natural Language

太田 吉美^{*1} 杉山 光裕^{*2} 村上 忠良^{*2} 大庭 英雄^{*3} 村田 晃一郎^{*4}
 Ota Yoshimi Sugiyama Mitsuhiro Murakami Tadayoshi Ohba Hideo Murata Koichiro

^{*1} 技術知識基盤構築機構

Institute of Technology Knowledge Infrastructure

^{*2} ライゼルトツシュ創知研究所

Riseltosh Research Institute

^{*3} 知的社会システム研究所

Emergent Synthesis Social System Agency

^{*4} 北里大学病院

Kitasato University Hospital

In order to realize healthy, safe and high quality of life, control of medical expenses and reservation of medical quality and safety are fundamentally important. For that purpose, we have to systematize the information and knowledge about medical treatment logically. It is indispensable to develop the Medical Knowledge Infrastructure which accumulates, shares and practical use of the information and knowledge about medical treatment. As such the Technology Knowledge Infrastructure, We developed the Formal Natural Language which is the semantic model of "subject + predicate + object", and describes the contents by ontology. We applied such technology to the medical field, and examined the framework for developing the Medical Knowledge Infrastructure. By development of the Medical Knowledge Infrastructure, medical treatment of a patient basis can be realized and medical increase in efficiency and improvement in quality and safety are expectable. Furthermore, it makes EBM possible and also realizes medical continuous improvement (PDCA cycle) based on risk management.

1. 緒言

健康で安全・安心で豊かな暮らしを実現するためには、少子高齢化対策や患者本位の医療の実現、医療費の抑制、医療の質・安全の確保などが基本的に重要な課題である。

医療分野でも、医療の効率化や質・安全の向上などのため、コンピュータを利用した医療の業務支援が行われており、いろいろなシステムが開発され利用されてきている。しかしながら、従来は都合がよく、効果の大きいところから情報システムの開発や導入を推進してきた。この結果、情報システムがばらばらで、孤島の状態になってしまった。これを解決するため、各種の標準化 (ISO/TC215, ICD-10, HL7, MML, CLAIM, DICOM 等) が行われているが、従来型の情報技術をベースにするもので、システムの柔軟性や将来への拡張性において多くの課題を抱えている。

これらの課題を解決するためには、医療に関する技術や業務の論理的な体系化および医療に関する情報や知識を蓄積・共有・創発できる医療知識基盤の構築が必要不可欠である。

筆者らは、このような技術体系として、製品や技術、サービスなどのライフサイクルに互い、情報や知識を蓄積・共有・創発するための技術知識基盤 (Technology Knowledge Infrastructure, 略して TECHNO-INFRA と呼ぶ) を構築し、そこに蓄積される各種指標をベースに、ビジネス・プロセスを継続的に改革 (PDCA サイクル, Plan, Do, Check, Action) することを提案してきた。技術知識基盤の基本技術として、「主語 + 述語 + 目的語」の基本構造 (セマンティック・モデル) で、各品詞の取り得る内容をオントロジーで限定する制約自然言語 FNL (Formal Natural Language) の開発を進めてきた [太田 2005, 2006]。また、制約自然言語に基づいた情報や知識の蓄積技術、情報や知識の共有・活用技術、既存のシステムやアプリケーション群との連携技

術などを開発してきた。

FNL は、人と人のコミュニケーションの手段として利用できるのは勿論であるが、コンピュータにも理解できる言語で、コンピュータ間や人とコンピュータとのコラボレーションとしても利用できる。

FNL では、オントロジーを拡張・拡充することにより、情報や知識を体系的に「積み上げる」ことができ、システムの柔軟性や将来への拡張性を飛躍的に高めることができる。

技術知識基盤のベースとなる、知識を蓄積・共有・活用するための FNL 及びその実装手段は、あらゆる産業や技術分野に適用でき、将来への変更や拡張が容易で、柔軟性のあるものである。これは、一般の製品や技術・サービスなどのライフサイクル支援システムが抱える多くの問題を解決できる。

本論文では、特に、医療分野への技術知識基盤の適用について報告する。

2. 技術知識基盤とは

あらゆる産業分野で、業務の効率化や製品の質・安全の向上などのため、コンピュータを利用した業務支援が広範囲に行なわれている。

しかしながら、従来は都合がよく、効果の大きいところから情報システムの開発や導入を推進してきた。この結果、情報システムが孤島になってしまった。情報の孤立や情報連携のミスが効率化の阻害要因であり、大事故や失敗の原因ともなる。業務の効率化や大事故・失敗を防ぐためには、あらゆる (網羅した) 情報や知識を統合し、これを論理的に活用できる環境の構築が必要不可欠である。

これからは、まず、情報システムの機能と情報を分離する。次に、分離した情報を統合化し、共有化する環境を構築する。さらに、これをグローバル標準にしてしまう。すなわち、情報のオーナーとしての立場を明確にする。製品や技術、サービスなどのライフサイクルに互い、情報や知識をコンピュータのハードやソフトウェアに依存することなく、蓄積・活用できる技術知識基盤を

連絡先：太田吉美，技術知識基盤構築機構，
 茨城県水戸市千波町 613-9，TEL:029-244-0875，
 yoshimi-ota@k3.dion.ne.jp

構築する。こうすることにより、コンピュータのハードメーカやソフトウェアメーカは、技術知識基盤と連携せざるを得なくなる。さらに、このような基盤が整備されると、新規参入が起りやすくなる。この結果、安く、性能のよい最適なシステムを選択できることにもなる。

また、このことは IT の急劇な進展による陳腐化などにも対応できることを示している。自らのコア・コンピタンスとなるような知的資産を、情報システムに依存しない形式で、長期にわたり管理できるようになる。

さらに、統合化した情報や知識をコンピュータが理解できるようにすれば、自動化の範囲が拡大できるし、人とコンピュータの協働作業支援を大幅に向上できる。蓄積した情報や知識から製品や技術・サービスなどの質や安全の向上、類似災害の防止など多くの活用が図れる。

このような技術知識基盤を構築するためには、製品や技術・サービスなどについての論理的な体系化及び知識の蓄積・共有・活用の機構化が必要不可欠である。また、技術知識基盤のベースとなる情報や知識を記述するモデル及びその実装手段（データウェアハウス技術、エージェント技術）を開発することが必要となる。

筆者らは、技術知識基盤のベース技術である、知識記述言語である FNL やこれに基づいたデータウェアハウス技術を開発してきた。また、技術知識基盤として、既存システム・アプリケーション群、情報や知識の蓄積層、情報や知識の共有層、情報や知識の活用層からなる4階層の基本アーキテクチャーを確立してきた(図1参照)。

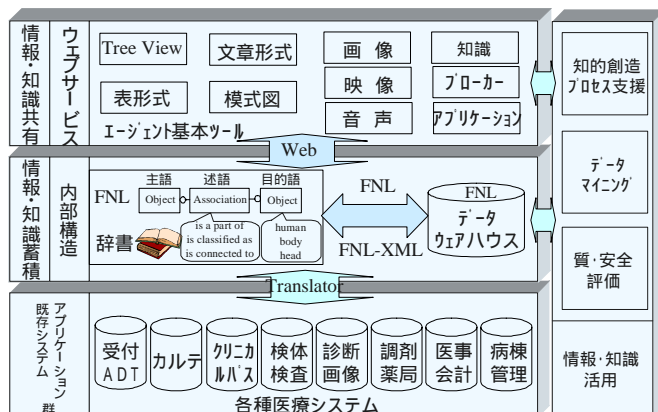


図1 技術知識基盤の基本アーキテクチャー

技術知識基盤を医療分野に適用した場合のフレームワークを図2に示す。医療知識基盤を中核として、医療や健康プロモーションなどの各プロセスで情報や知識の共有が行われる。これらのプロセスを担うものとして、国民・患者、医療を担当する病院・薬局、健康をプロモーションする検診センター、地域医療を担当する地方自治体、国全体を対象とする国や公共機関、医療を教育・研究する機関、医療や医薬に関連する企業など多くの機関が関与する。

医療知識基盤では、医療やサービスの各プロセスで必要となる情報を一元的に管理し、プロセスに関係する人がだれでも、いつでも、どこでも必要な情報を最適な表現形式で共有・活用することができる(図3参照)。ユーザーID やパスワードなどでセキュリティ管理されていることは言うまでもない。

以下では、技術知識基盤のベース技術である、制約自然言語 FNL や技術知識基盤の医療分野への適用について説明する。

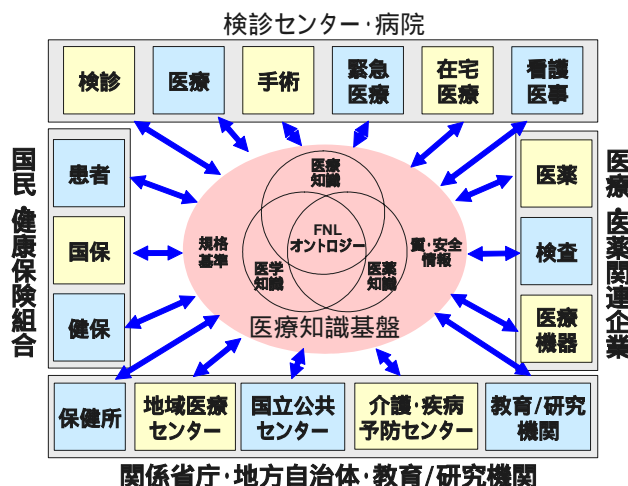


図2 医療知識基盤分野のフレームワーク

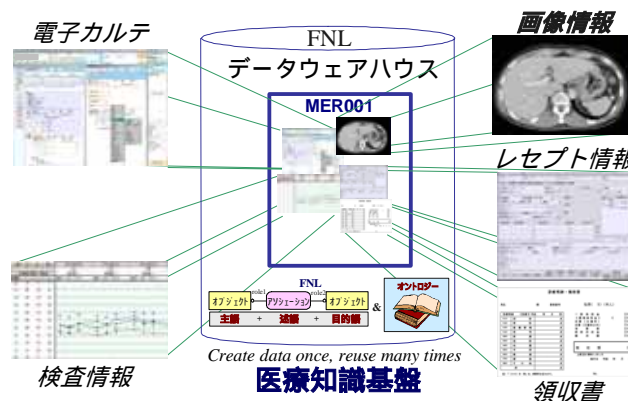


図3 データウェアハウスによる情報の蓄積

3. 制約自然言語 FNL

情報や知識を共有・蓄積するために必要なことは、情報や知識を表す用語(共通言語)を明確にすることである。このようなことは、ISO や IEC などの国際標準でも認識されている。FNL では、さらに積極的に、このような用語をオントロジーとして規定することを提案している。

FNL では、対象となる製品や技術・サービスのライフサイクル全般に亘り、ビジネス・プロセスを分析し(各プロセスの入力情報や出力情報を抽出)、情報モデルを構築し、情報や知識の共有に必要な概念(概念を規定するものをクラスと呼ぶ)や概念間の関連、概念と実体(インスタンス)間の関連、実体(インスタンス)間の関連(これらの関連の総称をアソシエーションと呼ぶ)を規定する。このため、FNL では次の四段階の開発方法を採用している(図4参照)。

(1) FNL コアモデル

FNL は、いろいろなデータ構造の基本要素(オブジェクトとオブジェクトの関連(アソシエーション))という要素を抽出したもので、この基本要素を組み合わせることにより、複雑な現実世界を記述しようとするものである。また、FNL は述語(動詞)を中心にした、意味ネットワークモデルであり、自然言語の構文(ステートメント)も記述できる重要な特徴がある。

FNL は、あらゆる製品や技術・サービスなどのライフサイクルに亘り、情報や知識を蓄積・共有・活用・創出することを支援する知識モデルで、従来のデータ構造やデータベースなどの考え方を革新する技術でもある。

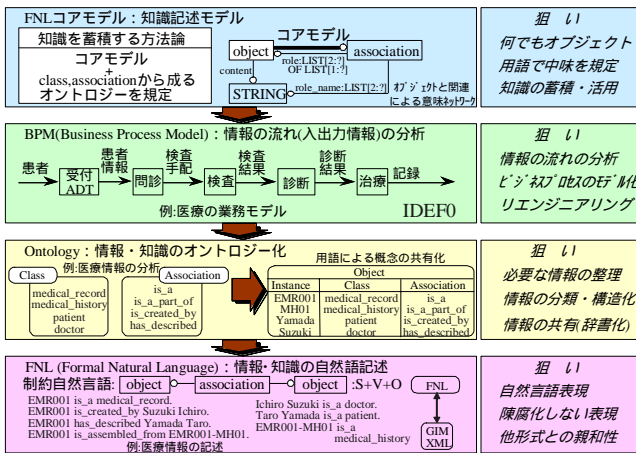


図4 制約自然言語 FNL の方法論

FNL では対象とするものすべてをオブジェクトとして捉える。オブジェクトは、概念(クラス)や実体(インスタンス)など定義対象の全てを記述するための入れ物(器)である。アソシエーションもオブジェクトの一つ(オブジェクトのサブタイプ)であり、オブジェクト間の関連付けを記述する器で、アソシエーションからオブジェクトには多数の手(ポインタ)を出すことができる。各々手にはそれぞれの役割(role)があり、その役割の内容についても明示的に記述する(図5参照)。

クラスとアソシエーションの中身は、後で述べるオントロジーで規定する。

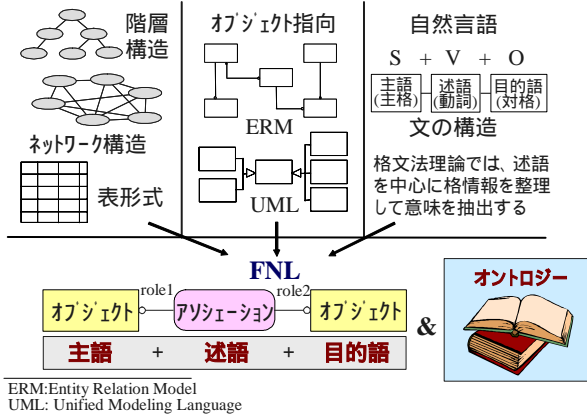


図5 制約自然言語 FNL の原理

(2) ビジネス・プロセスの分析

対象となる分野の業務内容や共有する情報や知識について、関係者の間で共通の概念、理解を持つことが必要である。このため、FNL では、ビジネス・プロセスの分析を行う。対象とする分野のビジネス・プロセスの分析を行い、各プロセスへの入力情報や出力情報などを明確にする。

FNL では、ビジネス・プロセスの各プロセスも一つのオブジェクトとして規定する。すなわち、各プロセスも一つのクラスとして規定する。これにより、各プロセス単位の情報や知識をまとめて表現することができる。

(3) オントロジーの整備

このステップでは、ビジネス・プロセスで分析した各プロセスでの入出力情報をまとめ、対象とする分野での情報や知識についての知識モデルを作成し、これをオントロジーとしてまとめる。

オントロジーは、さらに概念を規定するクラス、関連(述語)を規定するアソシエーションに分類する。

クラス定義では、概念を表す用語及びその意味を規定する。概念の意味の記述も制約自然言語を用いる。また、概念間の関連もアソシエーションを用いて記述する。

アソシエーション定義では、関連を表す概念の用語及びその意味(使い方)を規定する。アソシエーションでは、クラス間の関連、クラスとインスタンス間の関連、インスタンス間の関連などを定義する。また、アソシエーションは二項関係だけでなく、多項関係も記述できる。このため、アソシエーション定義では、多項関係の詳細を表す役割(ROLE)も規定し、アソシエーション用語と ROLE との関係を文脈(RULE)として記述する。また、このアソシエーション用語と同義語(ALIAS)を表すアソシエーション用語や、逆の関連(REVERSE)を表すアソシエーション用語も記述できる。

(4) FNL による情報・知識の共有

FNL では、概念の空間であるオントロジーと実体(インスタンス)を記述する空間とを組み合わせて表現する。これにより実際の情報や知識の体系的な管理が可能になる。

FNL には自然言語の構文も記述できるので、図6のように、主語、述語、目的語の順につなげると一つの文章となる。文章や文書を組合せることにより、知識を記述する。

FNL の基本構造は、主語 + 述語 + 目的語で、主語や述語、目的語の取り得る内容で、概念に関するものと述語(アソシエーション)はオントロジーで制限する。これにより、一般の自然言語処理の難しさを排除する。オントロジーは、英語、日本語、中国語、韓国語などの多言語に対応できる。

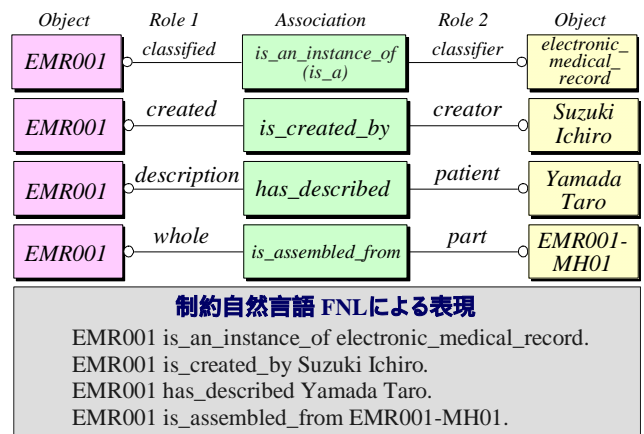


図6 制約自然言語 FNL の例

FNL は、オントロジーやインスタンス(実体で固有名称や具体的な数値を表す)の外部表現形式である。

FNL では、オントロジーであるクラスやアソシエーション定義から自動的にコンピュータが解釈する内部表現形式に変換される仕掛けになっている。FNL の外部表現形式では、同一の主語が何度も表れるが、コンピュータの内部表現では、同一の主語を持つ文をまとめて一つの複合した文として処理することにより、コンピュータの処理を高速化している。この変換により、用語の二重定義や用語間の論理的矛盾をパーサーでチェックする。

実体(インスタンス)は、各種情報システムのトランスレータを介して、コンピュータの内部表現として出力または、入力される。コンピュータの内部表現は外部表現形式である FNL に変換して、ユーザーに提示する。外部表現形式としては、FNL 表現以外にも、表や図形などに変換して表示することもできる。インスタンスに対してもパーサーにより論理的矛盾をチェックする。

4. 医療分野への適用

技術知識基盤の基本アーキテクチャーに基づき、FNL 及び FNL データウェアハウスなどの技術を医療分野に適用することを検討した。

医療分野の情報システムである電子カルテシステムや検査システム、画像システム、医事会計システムなどのデータベースの内容と FNL のオントロジーを参照し、マッピングテーブル(対応関係を記述する表)を作成する。この場合、データベースにある内容が FNL のオントロジーにない場合は、FNL のオントロジーに新しい用語として追加する。できあがったマッピングテーブルに基づき、各データベースとデータウェアハウスとの FNL トランスレータを開発する。FNL トランスレータにより、各種情報を既存のシステムから FNL データウェアハウスに登録し、FNL による表現に統一し蓄積する(図7参照)。

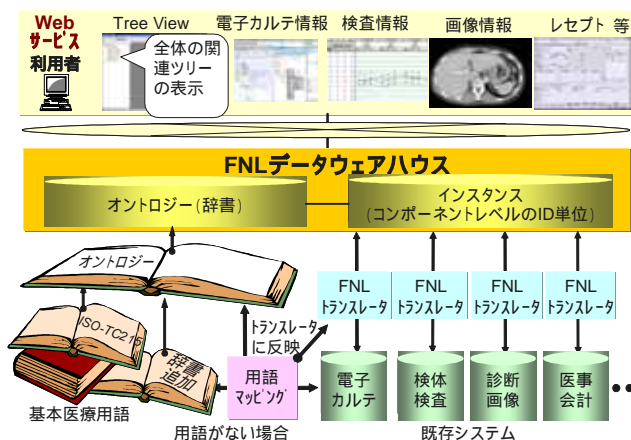


図7 制約自然言語 FNL の実装方法

データウェアハウス内の FNL で統一された内容は、Web 上でいろいろな視点から論理検索でき、表や図形、画像表示、自然言語的表示などの機能により閲覧できる。表示された内容は、コンポーネント単位で指示(ピック)ができ、関連する情報を検索・閲覧、加工することができ、Web 上での応用範囲が飛躍的に拡大する。

図8はデータウェアハウス内に蓄積した各種情報を、インターネットを介してデータウェアハウス内の情報を Web サービスであるエージェント基本ツールで検索・閲覧したイメージを示したものである。オントロジーで規定したクラスを表示(1)し、これをベースとして、各プロセスに互る情報や知識を検索・閲覧できる。例えば、(1)から対象となるカルテをピックアップして、カルテ情報を表示(2)し、次に、検査結果を表示(3)し、さらに、診断画像を表示(4)し、また、(1)から関連するレセプト情報を確認すること(5)が容易にできる。

医療知識基盤では、電子カルテシステムや検査システム、画像システム、医事会計システムなどの各プロセスに互り、すべての要素についての情報や知識を網羅的に統合化できる。

医療知識基盤の構築により、医療に関する情報や知識を一元的に蓄積、共有・活用でき、医療活動の効率化を図ることができる。また、各種の評価指標が蓄積できることにより、医療活動を継続的に改革する(PDCA サイクル)ことができ、さらなる効率化や医療の質・安全の大幅な向上が期待できる。

医療に関するすべての要素について、客観的で透明な尺度で一元的なマネジメントが可能であり、病院内は勿論、外部との

情報共有や連携にも有効である。また、医療知識基盤の構築により、以下の定性的効果を挙げることができる。

- (a)情報・知識共有による患者本位の医療の実現
- (b)医療業務の効率向上(業務のシームレス化)
- (c)医療の質・安全の大幅な向上
(情報連携の抜けミス防止)
- (d)リスク・マネジメントの徹底
- (e)医療活動の透明化
- (f)医療技術の伝承や新たな創発など

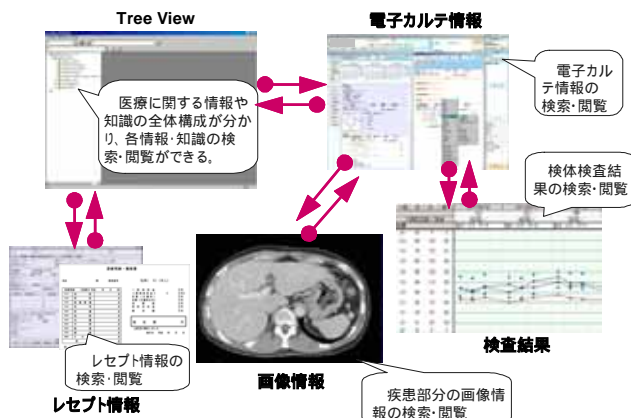


図8 医療知識基盤による情報共有

5. 結 言

医療知識基盤の構築により、医療業務をシームレスにつなげることが可能となり、情報の重複入力やミス、業務間で情報の欠落などを防ぐことができ、作業の効率化や作業期間の短縮、これらを通じた大幅なコスト低減などが期待できる。また、医療情報の網羅的な管理やエビデンスベースの医療(EBM)が可能となり、医療の質や安全の飛躍的な向上を期待できる。さらに、医療事例や故障・トラブルの事例を蓄積することにより、各種管理指標の精度向上を図ることができ、リスク・マネジメントに基づいた医療管理の継続的な改革(PDCA サイクル)が可能となる。

また、患者も含めた関係者間で医療の各プロセスに互る情報や知識を体系的に共有できることにより、日常の医療活動が効率化されるだけでなく、事故・トラブルに際しても、関係者間での一貫した対応やその迅速化、水平展開、原因究明の促進などにも役立ち、患者中心の医療が実現できる。

医療知識基盤の導入は、現在分断して行われている医療の関連業務に共通の価値観を与え、部分最適化の域を出ない現在の効率化を全体的な最適化に導くと共に、専門家も含め、専門家でない患者や関係者にも理解できる形で医療の運用実態を提示する透明性のある方法としても期待できる。

今後の緊急の課題としては、医療分野のオントロジー化を進め、医療知識基盤を実際に構築することである。医療分野では、既に、UMLS に見られるようにオントロジー化のベースとなるものがあるので、困難な課題ではない。

参考文献

[太田 2005]太田,大庭: 制約自然言語の実装方式, The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 2005
 [太田 2006]太田,好永,仲,芝尾: 技術知識基盤の構築 プロセス・プラント分野の情報統合化, The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006