

# デバイスオントロジーに基づく特許文書の意味理解モデルとその実現 特許を理解する計算機

## Semantic understanding model for a patent document based on device ontology

高藤淳\*1  
Sunao Takafuji

來村徳信\*1  
Yoshinobu Kitamura

溝口理一郎\*1  
Riichiro Mizoguchi

\*1 大阪大学 産業科学研究所

The Institute of Science and Industrial Research, Osaka University

Although intellectual property, especially patent, is getting more and more important so that wide variety of management methods and tools have been developed ever before such as search engines and patent maps, we yet don't have enough benefits from them. This is mainly because it is still very difficult for a computer to semantically grasp and process contents of a patent, even if computer science such as natural language processing has highly advanced in this age. When we rethink about a patent from the viewpoint of the description of the structure and functionality of an artifact invention, we realize that the description implicitly includes functional ontology and can be interpreted by the ontology. This paper describes new semantic understanding model for a patent with using device ontology as cognitive schema.

### 1. はじめに

近年、知財管理、特に戦略的視点からの特許管理の重要性がより鮮明になり、特許性の早期確認、出願リードタイムの短縮、保持・廃棄の意志決定、国際出願対応等が企業のみならず国家的見地からの課題ともなっている。このため特許検索やパテントマップをはじめとする IT による知財活用基盤が整備され利用されているが、未だ特許の意味内容に踏み込んだ高度な IT 支援は実現しておらず、知財関係者は大きな特許理解のコスト負担を余儀なくされている。従って、特許の意味内容に基づく計算機支援が実現すれば特許戦略に少なからぬ貢献をもたらすことが期待されるが、対象の広範囲性や高度な専門性、文章表現の特殊性などにより人間と等価な深層レベルの意味理解を計算機処理で実現するのは困難な状況にある。

ここで特許文書の内容を鑑みると、「発明は自然法則を利用した技術的思想の高度な創作である」ことから、必然の帰結として何らかの科学的原理に依拠する装置類の機能、構成、仕組み・方法等が詳述されており、デバイスオントロジー[來村 02]に基づいた解釈が可能であることが示唆される。実際、ナノテクノロジー材料合成プロセスに関する機能的知識体系化の研究[垂見 03]においては、特許文書から人が原初的な機能的知識を読み取れることが指摘されている。

一方、人の代わりに計算機が特許の文章を読み取り解釈する処理を実現するには読解プロセスのモデル化が必要である。認知心理学研究において読解プロセスはボトムアップ処理(データ駆動処理)とトップダウン処理(概念駆動処理)によって構成され、読み手がテキストを再構成する過程と捉えるのが一般的である。例えば、スキーマ理論[Carrell 83]では読み手が持つ背景知識をスキーマと位置づけ、テキストから情報を得てスキーマにマッピングするボトムアップ処理と、スキーマによるテキスト情報の採択というトップダウン処理が同時並行的に相互作用して読解が行われるとしている。この理論では相互作用の具体性が十分ではないが、構成-統合モデル[Kintsch 88]では、語彙のネットワークを構築するボトムアップ的な構成プロセスと理解の一貫性を与えるトップダウン的な統合プロセスを語彙ネットワークにおける活性伝搬を用いて実現する具体性のあるコネクショニ

スタプローチを提案している。構成-統合モデルは一般的な読解メカニズムを説明するものであるが、世界知識を前提としており現実的ではない。

そこで、筆者らは計算機による実用的な特許文書読解支援処理を指向して、人工物の把握にとって重要な機能概念以外の意味論を極小化し、デバイスオントロジーを認知スキーマとしてスキーマ理論を拡張した上で文脈や意図を包括する理解のメンタルモデルに基づいて、スキーマ要素の充足を近似的な理解状態と見なす特許文書の意味理解モデルを考案した。

本稿では、次章においてデバイスオントロジーと機能分解木について述べた後、3章において筆者らが考える特許文書の意味理解モデルについて詳説する。さらに4章では本意味理解モデルを用いた特許文書読解支援システムのプロトタイプを紹介し、最後に5章でまとめと今後の課題を述べる。

### 2. デバイスオントロジーと機能分解木

#### 2.1 デバイスオントロジー

オントロジー工学では人工物の機能的知識に関わるオントロジー・知識の階層を図1のように体系化している。

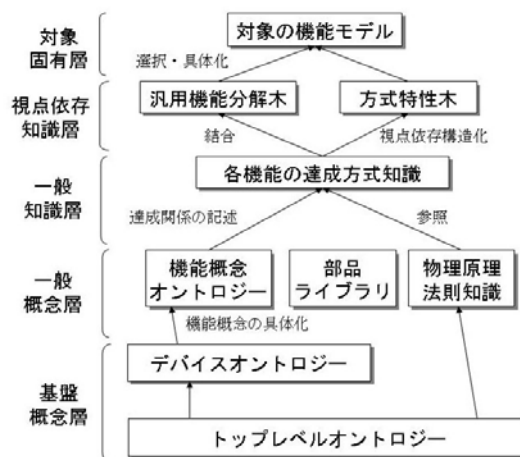


図1. オントロジー・知識の階層

連絡先: 高藤淳

大阪大学産業科学研究所 新産業創造物質基盤センター

〒565-0047 大阪府茨木市美穂ヶ丘 8-1

Tel: 06-6879-8424-5050 Fax: 06-6879-8424-5050

e-mail: takaj@sanken.osaka-u.ac.jp

デバイスオントロジーは基盤概念層に位置してデバイスを中心とした視点によって人工物の機能をモデリングするための基盤となるオントロジーであり、従来の概念である入出力ポートに加えて、処理主体となる「装置」、単なる伝達のみを行う特殊な装置である「導管」、装置の処理対象である「対象物」、対象を担う「媒体」の4つの概念を導入することによって人工物を一貫性のある形式で表出化する規範を与える。

本研究では、デバイスオントロジーが本来的に有する普遍性のある人工物の概念的形式化能力を人が人工物を理解する際に内的に保持する認知スキーマと等価に見なせる点に着眼して、後節において認知科学的視点からの意味理解モデルを考察する。

## 2.2 機能分解木

機能分解木はデバイスオントロジーに立脚し全体装置(部品)から部分装置(部品)へ「機能」(AND結合)と「方式」(OR結合)を交互に機能分解した木構造によって人工物の機能的知識を表出化するものであり、機能や方式に付随する様々な情報を同時に網羅的に明示化する機能性を提供する。

機能分解木を記述することによって背景の異なる関係者の知識共有が容易となることから、デザインレビューの円滑化、不具合診断の迅速化などの他、特許文書申請処理時間の短縮に関する効果が報告されている[溝口 02]。特許文書理解の文脈においては、技術者と弁理士が短時間に理解を共有することができたとされている点が特に重要であり、これは機能分解木を通して両者の人工物理解のためのメンタルモデルが極めて近いものとなっていたことを示唆している。

このような効果を近似的に示すため実際の特許文書[特開95]を簡易な機能分解木に書き起こした例を図2に示す。

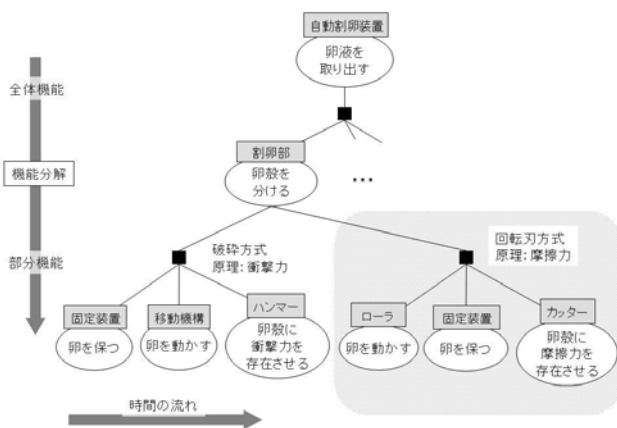


図2. 自動割卵装置の機能分解木

図中楕円は何を達成するかという「機能」を表し、小矩形は達成原理を概念化した「方式」を表しており、前述のように全体から部分へ機能と方式を交互に機能展開している。また、同一の部分機能列は左から右に時間的な流れを示している。本ケースの特許では背景がグレーの面取矩形領域が発明箇所を示しており、その左に記述された従来の技術、つまり「衝撃力による破碎方式」に対して、「摩擦力による回転刃方式」による自動割卵機を発明したということが方式の対比として明瞭に確認できる。

機能分解木は本来概念設計などの設計フェーズにおいて人が任意の人工物の機能的知識を明らかにするために記述するものであるが、本研究では計算機が人工物に関する記述を解釈して「分かった」後の状態を現す表出化形式として援用する。

## 3. 特許文書の意味理解モデル

本章ではスキーマ理論を発展させた特許文書の意味理解モデルを検討する前提として、まず人が何かについて「分かる」とはどのようなことであるかについて関連分野の知見を加えた上で再考し、次に「読んで分かる=読解」とはどのようなことを意味するかという点について考察を深めた後、筆者らが考える意味理解モデルの定式化を試みる。

### 3.1 「分かる」ということ

「分かる」ことについて考える場合、「分かる」対象と利用可能な情報や知識、そして処理過程が中核的要件と考えられる。まず、「分かる」対象に関しては、五感で知覚できる存在物、内的心象として構成される抽象概念、観察と思考によって到達可能な物理現象の背後に潜在する自然科学の法則など異なる特性を持つ諸処の対象が存在する。また、利用可能な情報とは視覚であれば眼球の解像度で捉えられる視野内の画像情報であり、聴覚であればある距離内での可聴周波数域の音声信号である。また、生物学的な知覚情報に意味を付与するためには記憶に基づく知識が必要である。そして最後の処理過程は脳内における意味体系の統合プロセスに関わるものであり、対象や知覚情報によって複合的である。

最近の脳科学は脳イメージング技術の発達により前記五感の知覚情報が脳内でどの領域がどのように賦活されているか、或いは予測的理解を促すプライミング記憶、海馬と長期記憶の関与性などの記憶の種類や所在を観測可能としている。これらは高次機能障害すなわち「分からない」ことを可観測にできることと等価であり、例えば、物理的な視覚機能が正常に機能していてもその意味が分からないことが何故生じているかという双対問題によって「分かる」ことへの理解へアプローチしている。

このようにメタ認知的な脳の高次機能について個別の局所的な解明は進みつつあるため、統合的な意味理解プロセスは依然未知の状況であっても、脳科学等の知見を援用することによって心理学実験から導出されてきた旧来のスキーマ理論等の仮説を修正・拡張しつつ計算機処理用の新たなモデルに洗練できると考えられる。

ここで一旦、「分かる」ということは一般に脳の複合的な相互作用であり、知覚情報を脳内で統合的知覚パターンとして合成し、複数の記憶メカニズムに基づく記憶情報に照らし合わせて対象に対する新たな意味の体系を心象として形成することとまとめておく。

### 3.2 「読んで分かる」ということ

本節では「読んで分かる」ことを子細に検討する。「読んで分かる」ことは前節のまとめを特殊化して、対象が二次元平面上の言語記号列であり、視覚を通して言語記号列を認知しつつ各種の記憶を参照して意味の体系を内的に形成することと表現することができる。視覚による言語情報の認知の問題はスキーマ理論のボトムアップ処理における読解の起点となるインプット情報をより具体化する問題と重なる。

認知科学における物体の視覚認知の研究では対象物に対する注意は一樣ではなく焦点的なものと分散的なものに区分される[齋木 05]。つまり、生物学的な眼球の視野に対して、脳は情報視野とでもいうべき認知的視野を内的に保持し、注意によって先鋭化して認識の度合いが高まることを意味する。

読解における視覚認知の対象は二次元平面上の線形な言語記号列であるが、物体の場合と同様に情報視野が機能して、焦点化された注意は基本的に一方向の直線的移動によって

「読み進む」プロセスを進行させる。短期記憶の保持が困難な場合等は「読み戻す」処理を行う。「読み進む」過程において意味を認識し得るある単位の言語情報の断片(以下、言語断片)が読み手の注意の特徴点となり、この注意の移動と同時に意味を生起する。特徴点の言語断片の意味はスキーマによって同定されるが、これはスキーマ理論における同時並列的なトップダウン処理に相当する。また、注意は認識と結びついているため、視覚の認識領域が分化していることを踏まえると、注意にも次のような複数の種別が存在すると仮定できる。

- 該当言語記号列固有の構造的特徴に対する注意
- 事物・事象に対する注意
- 事物・事象の関係に対する注意
- 事物の機能発揮に対する注意

そこで、このような情報視野の情報処理モデルとして情報視覚モデルという仮想的なモデルを導入する。情報視覚は固有の注意に基づいて特定の言語断片のみを認識し、かつ動的な焦点移動を司る複数の情報レンズからなり、認知した個々の知覚情報は作業記憶に保持する。読解プロセスはボトムアップ処理に相当するこの情報視覚による処理と期待駆動型の認知スキーマによるトップダウン処理が同時的な相互作用として機能することで進行する(図3)。

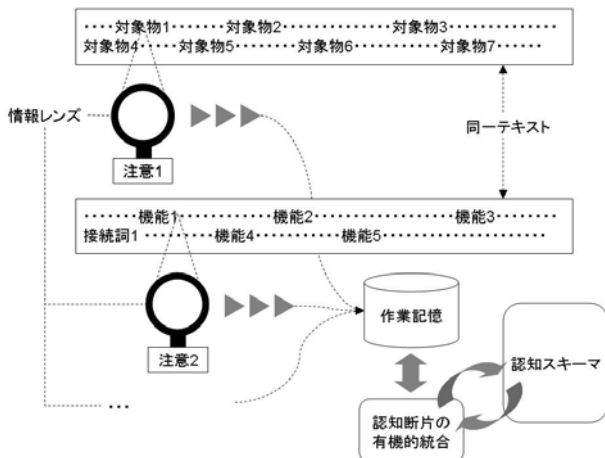


図3. 情報視覚モデルによる読解プロセス

### 3.3 認知スキーマとしてのデバイスオントロジー

本節では、スキーマの内部表現を詳細化し読解のメンタルモデルとして再構築するとともに、特許文書において何故デバイスオントロジーがスキーマとして適格であるかについて論ずる。

スキーマ理論では、スキーマをテキストの形式的な修辞構造に関する背景知識である形式的スキーマとテキスト内容の背景知識である内容スキーマに区分している。筆者らは形式的スキーマについて文書の論理構造を加えるとともに、内容スキーマについてオントロジー工学に基づく情報構造を導入した上で、文脈や読解の意図によってそれらを包含した総体を読解のメンタルモデルとして敷衍する。

まず、形式的スキーマにおける文書の論理構造の導入は、個々の局所的な修辞構造をマクロ的に規定し書き手の意図を背景に含むため重要なスキーマ要件の一つと考えられる。また、内容スキーマは例えば物語スキーマであれば物語の登場人物やイベント、プロセスによって構成されるが、これらはインスタンスとして存在するのではなく、一般化された概念体系として存在

するため概念と概念間の関係を定式化するオントロジーによって表現することは自然と考えられる。

認知スキーマは静的な構造の他に、動的な機能性も内包している。例えば、期待駆動によるトップダウン処理を担う機能であり、注意の焦点と呼応して期待するスキーマ要素を予測する。また、文脈によってスキーマ要素間の関係が変化する他、意図すなわち「要点のみ分かりたい」、「完全に分かりたい」など理解の目的に応じた制約もあるため制約機能が存在する。

以上から読解のメンタルモデルを模式化したものを下図に示す。

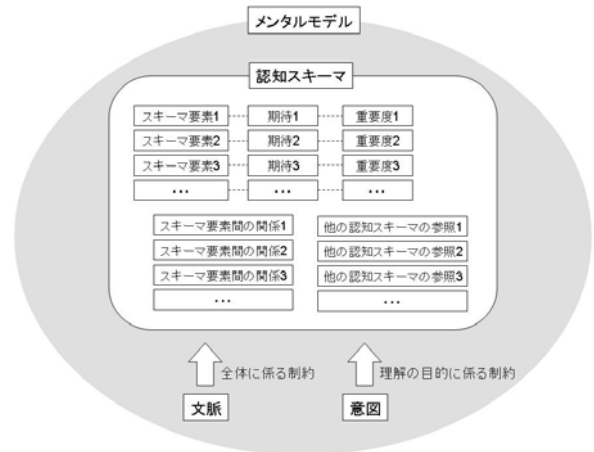


図4. 読解のメンタルモデル

認知スキーマは個々人の内的な表象として形成されるものであるが、人間同士のコミュニケーションの基盤となる共有性を重要な性質として保持するものでもある。特許文書の場合は対象となる人工物の認知スキーマが人々の間で共有可能な汎化された心象として潜在的に存在することを仮定できる。このような観点に基づくデバイスオントロジーは本来人工物一般に対する機能モデルの概念的規範としての表現様式を厳密に規定するものであるため、理解された状態の表現系として極めて適格性があるものと考えられることができる。

例えば、実際の特許文書[特開 99]の内容に基づいて前述の適格性を次のように確認することができる。

【請求項1】燃焼部と、この燃焼部での燃焼熱を受ける熱交換部と、この燃焼部に燃焼空気を供給するファンとを備えたガス燃焼装置において、熱交換部の下流側に温度センサを設置するとともに、検出手段を備え、上記検出手段は、上記燃焼部に供給されるガス量と、上記熱交換部で消費される熱量と、上記温度センサで検出される排気温度とに基づき、燃焼空気および排気ガスの通路の目詰まりを検出することを特徴とするガス燃焼装置。

本例では、全体装置であるガス燃焼装置が燃焼部や熱交換部、ファンなどの部分装置を持ち、燃焼空気の供給や目詰まりの検出等の機能を発揮していることを読み取れる。これらは前章で述べたデバイスオントロジーにおける処理主体、対象物、媒体等のオントロジー構成要件とほぼ一意に対応するものであり、それら構成要件に複数の注意を割り当てて文章全体をスキャンすると特許の対象となっている装置のデバイスオントロジーの要件を完全ではないにせよ相当程度充足できることが分かる。

### 3.4 特許文書の意味理解モデル

前節では情報視覚モデルによって特許に記載されている人工物の認知スキーマを人工的に構成できる可能性について述

べたが、実際には様々な問題がある。例えば、特許の文面通りの記述で用いられている機能語彙(以下、表層語彙)はオントロジー学的な機能概念(以下、標準語彙)を表すものではない。このため表層語彙を標準語彙にマッピングするための枠組み[高藤 07]を用いて機能概念を同定する機能性が必要となる。また、自然言語の記述上欠落する認知スキーマの情報を既知の機能的知識によって補完する機能性も必要である。以上を踏まえて、最終的に機能分解木として統合するプロセスを完備した総合的な意味理解モデルを図5のように構想した。

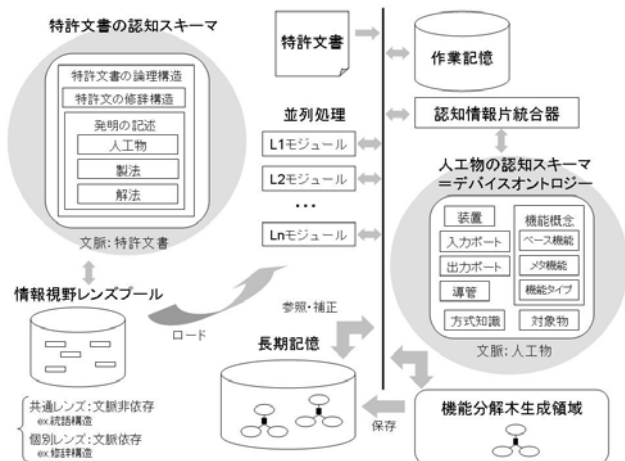


図5. 特許文書の意味理解モデル

この意味理解モデルは、読解の開始時点において異なる複数の情報レンズを保持する情報レンズプールから必要な情報レンズを動的にロードすることで内部的に情報レンズ解析モジュール(図中 Ln モジュール)を形成する。また、前記各解析モジュールの解析結果を統合する認知情報片統合器を持ち、解析時には解析結果が書き込まれる作業記憶をモニタリングしてデバイスオントロジーの構成要素を識別しながら、機能分解木の形式にまとめ上げる。構成要素が機能概念の場合は、前述した表層語彙から標準語彙へのマッピングにより機能の概念化を行い必要な対象物を予測的に同定する。機能分解木は個人にとっては人工物の機能的理解を促し、背景の異なる人間同士にとっては相互理解を促進し、さらには計算機間でも該当オブジェクトの交換により同一の解釈を可能とする人工物理解の可搬的な媒体であり、本モデルはその機能性を具現化するものである。

#### 4. 意味理解モデルに基づく特許文書読解支援

前章で考察した特許文書の意味理解モデルを用いることによって人間が特許文書を読解する際の支援を与えるシステムを実現することができる。筆者らは、本枠組みによる意味理解モデルの妥当性を検証する実験基盤として特許文書読解支援システムのプロトタイプの実装を試みた。

本プロトタイプは XML 形式の特許文書を対象として、指定した複数の特許文書を読み込み、選択した任意の請求項に対して意味理解処理を実行することで機能分解木を生成する。このプロトタイプでは情報レンズとして対象物の認識、機能概念の認識、機能間の関係の認識などを司る複数の有限状態オートマトンを持ち、これらが並列なボトムアップ処理により個々の機能分解木の要素とミクロな関係を同定する。一方、機能と方式の機能分解構造という機能分解木の基本フレームが充足すべき要素として何を期待するかというトップダウンの期待充足処理を同時に実行することで、全体の機能分解木構造が生成される。

ただし、現段階の実装では読解の支援には有意であるが、オントロジー工学的に妥当な機能分解木にはなっていないため、生成された機能分解木においてオントロジー工学的に不適切な箇所があれば、画面上で修正することで適正な機能分解木が構築できるようになっている(図6)。

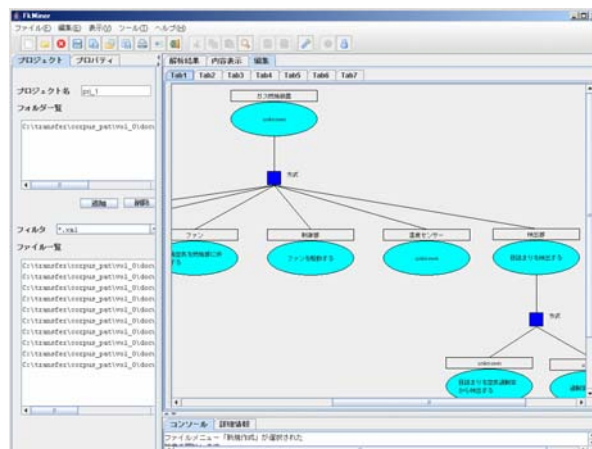


図6. 特許文書読解支援プロトタイプ

#### 5. まとめ

本研究では特許文に記述された人工物を理解するメンタルモデルの基底としてオントロジー工学におけるデバイスオントロジーと機能分解木を用い、動的な注意の複合的焦点移動を特徴とする情報視覚モデルの導入によりフレーム理論の拡張を行うことで計算機による特許文書の意味理解モデルを定式化した。

さらに、本モデルに基づく特許文書読解支援システムについてその概念プロトタイプを構築してモデルの実効性を確認した。

今後の課題としては、規範となる人工物の認知スキーマオーサリングや情報レンズの多様化・精緻化による機能分解木の精度向上、既存知識による認知フレーム補完機構の実現などがあげられる。

#### 参考文献

[來村 02] 來村徳信, 溝口理一郎: オントロジー工学に基づく機能的知識体系化の枠組み, 人工知能学会誌, 2002.  
 [垂見 03] 垂見晋也, 古崎晃司, 來村徳信, 渡邊英一, 溝口理一郎: ナノテクノロジー材料合成プロセスに関する機能的知識体系化の枠組み, 第 17 回人工知能学会全国大会, 1G2-04, 2003.  
 [Carrell, P.L. 83] Carrell P.L: Schema theory and ESL Reading pedagogy. TESOL Quarterly. Vol.17 No.4, 1983  
 [Kintsch 88] Walter Kintsch: The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model, Psychological Review, Vol. 95, No2., 1988  
 [溝口 02] 溝口理一郎, 來村徳信, 布施雅義: オントロジー工学の成功事例: 人工知能学会・知識ベースシステム研究会 (第 57 回), 2002  
 [特開 95] 日本国特許庁 卵自動割り方法及びその装置 特開平 7-289170  
 [齋木 05] 齋木潤: 視覚認知における知覚と記憶, Technical Report on Attention and Cognition No.14, 2005  
 [特開 99] 日本国特許庁 ガス燃焼装置 特開平 11-241820  
 [高藤 07] 高藤淳, 來村徳信, 溝口理一郎: オントロジー工学と自然言語処理の融合による機能分解木記述支援の枠組み, 電子情報通信学会, NLC 研究会, 2007