

# オントロジーと特許分類

大学共同利用機関法人情報・システム研究機構 国立情報学研究所学術コンテンツサービス研究開発センター長 **武田 英明**

## PROFILE

1991年3月 東京大学大学院工学系研究科 博士課程 修了。ノルウェー工科大学、奈良先端科学技術大学院大学を経て、2000年より国立情報学研究所助教授。2003年同所教授。2006年より同所学術コンテンツサービス研究開発センター長。2005年～2010年東京大学寄付部門教授（客員教授）。専門は人工知能、Web情報学。工学博士。

✉ takeda@nii.ac.jp

TEL 03-4212-2543

## 1 はじめに

特許情報は個別に産業に有用な情報であると同時に総体としては人類の知の集積でもある。知の集積として特許情報を見ることは特許情報に新しい視点を与え、その結果として新しい利用の仕方ができることが期待される。

その一つの視点がオントロジーである。本稿では特許分類とオントロジーの関係を論じ、22年度に行った簡単な分析について報告する。

## 2 知の体系化としてのオントロジー

オントロジーとは本来、哲学の存在論のことであるが、90年代より計算機科学ことに人工知能において計算機内に表現する存在物に関する記述を指すようになった。形式的には“世界にある物事に関する概念と概念間の関係の定義”である。それまでは知識はプログラムごとに個別に記述してきたが、オントロジーはプログラム共通の知識として利用することができる。

典型的なオントロジーは論理的方法で定義される。またオントロジーの多くは概念間に上位下位関係 (isa 関係) をもち、階層的な構造を形成している。他にも全体部分関係 (part-of 関係) を多く用いられる。さらに独自の概念間の関係も定義され使われる。

オントロジーはいくつかの階層に分けられる。トップ

レベルオントロジーとは世界の物事の存在のあり方を規定するもので、最上位の「もの」「こと」から始まって、抽象的な存在や具体的な存在を定義する。ドメインオントロジーとは各学問分野などにおいて使われる概念を体系化したものである。理論上はドメインオントロジーはトップレベルオントロジーのより具体的なものであるが、必ずしも整合的にできているわけではない。

## 3 オントロジーと特許分類

本章では、特許で用いられているカテゴリー (IPC 分類、FI、Fターム) をオントロジーとしてみたときに、どのような構造が見えてくるかについて考察する。先に述べたように一般的なオントロジーの構造は上位下位関係 (isa 関係)、全体部分関係 (part of 関係)、その他の関係で概念同士が関係づけられている。

特許カテゴリーは基本的に分類体系なので isa 関係であると考えることができる。しかし、Fタームは独自で観点というものを持つ。そこでFタームにおいては観点を独自の関係とみたときにどのようなになっているかを考察する。

### 3.1 IPC 分類

IPC 分類はセクション、クラス、サブクラス、メイングループ、サブグループというものから構成されている。またサブグループ間にはドット階層と呼ばれる階層構造がある場合がある。これらは基本的に上位下位関係であ

る。各特許はメイングループまたはサブグループに関係づけられる。

すなわちオントロジー構造としては、セクション、クラス、サブクラス、メイングループ、サブグループの要素はこの順に isa 関係で関係づけられる。またサブグループの要素は別のサブグループの要素と isa 関係があることもある。個別の特許はメイングループあるいはサブグループに関係づけられる。

原則的には木構造を成すが、一部他の枝の要素の関係づけがある。また、留意点が必要なのは、一定期間ごと改訂なので対応する分類が変わる点である。

### 3.2 FI

FI は IPC 分類の詳細化に当たる。IPC 分類記号の末尾に展開記号 (3 桁数字) あるいは分冊識別番号 (1 桁英字) あるいはその両方 (展開番号 + 分冊識別番号) をつけたものである。

オントロジー構造としては全ての FI の要素は IPC 分類のメイングループかサブグループの isa であり、FI の要素のうち、展開番号・分冊識別弁号付き FI 記号は展開番号付き FI 記号の isa である。

### 3.3 F ターム

F タームはテーマと呼ばれるものを用意して、個別のテーマを構造的に記述している。その際、観点というものが用意され、観点の下に個別の F タームが定義されている。例えば 2B003 「人工魚礁」というテーマについての記述であり、そこでは 5 つの観点が定義されている (設置形態、構造、構造以外の特徴、材質、対象生物)。観点 2B002AA00 「設置形態」の下には 4 つの F ターム (「・沈設型」「・浮沈型」「・浮き型」「・護岸」というものがある。ここでもドット階層が用いられていて、この例では「・浮沈型」は「・沈設型」の下位ということになる。

また、テーマは FI とは FI 適用範囲という形で、FI と関係づけられている。この例では、対応する FI は A01K61/00,311-61/00,321 である。

この状況をオントロジー的には次のように考えること

ができる。

まずテーマはテーマグループというものでまとめられている。またテーマは FI とは FI 適用範囲という形で、FI と関係づけられている。

## 4 F タームの統計的調査とからみるオントロジー化の可能性

以下では、この F タームについて、上記の構造がどうなっているかを調査した。今回は F ターム内における構造のみに着目して分析した。

なお、F タームの観点については [1] では 2 通りがあると述べている。

- (a) 発明の特徴点を単純に類型化した観点
- (b) 発明の特徴点を技術要素の組合せにより類型化した観点

2 つの違いは直交的であるかどうかであると書かれている。実際に特許の参照関係を大規模に分析するとこの 2 つは分かれる可能性があるが、今回は特許事例との関係は利用しなかった。

#### (1) タームの階層構造

全体では 2479 個のテーマがあった。また F タームは 363,774 個あった。

テーマでの平均を取ると、1 テーマあたり 146 個の F タームがあることになる。テーマがクラスだと考えると、これがクラスの平均的サイズということになる。この大きさはテーマグループごとに偏りがある。4F グループは最大で約 482、次が 4J となる。小さい方は 5L が最小で約 45、次が 2K で約 62 である。

1 テーマ毎の観点の数の平均は約 9.5。すなわち一つのテーマには概ね 10 個程度の観点が用意されていることになる。これもテーマグループ毎の偏りが大きい。最大は 4J で約 23、次は 4F で 22。最小は 5L で約 2、次は 3K で 3。概ね F ターム数と同様になっている。

もっとも観点毎の F ターム数は一定のばらつきがある。観点毎の F ターム数は平均で 15.4、グループ毎の平均は 11.0 ~ 22.1。しかし、テーマ毎の観点数や F ターム数の分布に比べれば小さいので、観点毎の F ター

ム数は比較的安定しているといえる。

ドット階層の構造はどうなっているだろうか。ドット階層は最大で8である。Fターム数で最大の階層は1階層目であるが、1階層目と2階層目はほぼ同数である。3階層目は2階層目の約1/3、4階層目もその1/4程度となる。概ね4階層目までである。一番末端になるFタームは全体の70%程度である。すなわち、30%ぐらいが下の階層をもつFタームとなる。グループ毎では0.68～0.75で、グループ毎の差異は小さい。平均的な深さは1.7。これはグループ毎では1.5～2.0で差異は小さい。

以上を通じて分かったのは、テーマグループによってテーマの大きさを大きく変わるが、それはテーマ毎の観点数の差異によるもので、観点の下のFターム数は概

ね安定している。またドット階層の構造はグループごとにそれほど差はない。

## (2) 観点の分析

次に観点にはどのようなものがあるかを調べた。図1は観点から名詞を取り出したものである。「装置」を含む観点は1853、「構造」を含む観点は1617、以下「制御」、「目的」、「手段」、「処理」、「構成」、「物」、「対象」、「部」と続く。

この分析では形態語解析をして単語を切り出しているため、それぞれが必ずしも単独で意味をもつ語とは限らない。例えば「物」「部」「もの」などは他の名詞の一部だったものが多く含まれると思われる。このため、それらは以下の分析では外して考える。

これらの名詞を含む観点を含むテーマ数を3列目に

名 詞	出現観点数	出現観点テーマ ユニーク数	全テーマに 占める割合
装 置	1853	742	0.299314
構 造	1617	749	0.302138
制 御	1510	676	0.272691
目 的	1133	916	0.369504
手 段	1133	551	0.222267
処 理	1117	553	0.223074
構 成	939	498	0.200887
物	849	520	0.209762
対 象	823	518	0.208955
部	809	798	0.321904
用 途	806	664	0.26785
特 徴	758	329	0.132715
種 類	744	504	0.203308
材 料	701	367	0.148044
その他	689	500	0.201694
機 能	653	468	0.188786
方 法	600	396	0.159742
検 出	572	309	0.124647
製 造	557	349	0.140783
も の	553	292	0.117789

図1 観点に出現する名詞の上位20種

示す。最大は「目的」で916個、以下「構造」749、「装置」742と続く。テーマ数が2479であったので、「目的」を含む観点は約1/3のテーマで使われているということが分かる。

この割合で20%を目安として選択し、また単独で意味をもつ名詞をみていくと、

構造／制御／目的／手段／処理／構成／対象／用途／特徴／種類

の10個が残る。これらは観点記述の基本的な用語であるといえる。

テーマを観点を属性とするクラスと考えるならば、この10個はテンプレートとして使える属性群とみることができる。

## 5 まとめ

今回は特許に関わる分類カテゴリーをオントロジー的視点から分析を行った。ほとんどの構造は上位下位関係であるが、Fタームには明示的ではないが非上位下位関係があり、暗黙のクラス構造があることが分かった。

ここででてくる属性はカテゴリーを超えて共通の属性であることと思われる。これらの属性は単にカテゴリー構造の分析だけでなく、特許自身の記述にも関係していると考えられる。この点についてはさらに分析することが必要であろう。

### 参考文献

[1] 国際特許分類、F1、Fタームの概要とそれらを用いた先行技術調査、平成20年度知的財産権制度説明会(実務者向け)テキスト、特許庁(2010)、[http://www.jpo.go.jp/torikumi/ibento/text/pdf/h20\\_jitsumusya\\_txt/02ipc\\_01.pdf](http://www.jpo.go.jp/torikumi/ibento/text/pdf/h20_jitsumusya_txt/02ipc_01.pdf)