

分類検索システムの一般設計論の必要性

Necessity of General Design Theory of Classification Retrieval System

著者：上田 育弘

Ikuhiro Ueda

ベストフレンド知的財産研究所

¹Best Friend Intellectual Property Laboratory

Abstract: Necessity of General Design Theory of Classification Retrieval System is discussed, comparing Classification retrieval system (ex. Fterm retrieval system) and Keyword retrieval system (ex. Google, Yahoo), in order to search for the most necessary information rapidly and correctly in the internet society. Currently there is Fterm retrieval system in which the classifications are given beforehand, differently from Keyword retrieval system. The characteristic of this Fterm retrieval system lies in excluding the ambiguity of the keywords. Meanwhile the characteristic of Keyword retrieval system such as google or yahoo lies in thoroughly excluding giving the classifications beforehand. So Classification retrieval system (ex. Fterm retrieval system) and Keyword retrieval system (ex. Google, Yahoo) have utterly different design theory in ‘giving the classifications beforehand’ and ‘keyword.’ Especially the possibility for applying graph-theory is discussed in Necessity of General Design Theory of Classification Retrieval System in this paper.

1. 問題提起（概要）

分類検索システムとしてFターム検索システムとグーグルやヤフー等のキーワード検索システムとの比較異同を通して欲しい情報を的確かつ迅速に探し出す分類検索システムの一般設計論の必要性を論じる。インターネット社会の到来によりインターネットを介して欲しい情報を的確かつ迅速に探し出す必要性が随所で顕在化している。現在、事前に分類を付与する分類付与型の検索システムとして特許情報のFターム検索システムが存在している。このFターム検索システムの特徴は事前に分類を付与することに徹しており、いわゆるキーワード検索によるキーワードの多義性を徹底的に排斥している点に大きな特徴がある。一方、グーグルやヤフー等のインターネット検索システムの特徴は事前に分類を付与することを徹底的に排斥しており、キーワード検索に固執している点に大きな特徴がある。従って、Fターム検索システムとグーグルやヤフー等のインターネットにおけるキーワード検索システムとは、「事前に分類付与」「キーワード検索」に関し全く正反対の設計思想を有している。このように「事前に分類付与」「キーワード検索」に関し全く正反対の設計思想を有しているFターム検索システムとグーグ

ルやヤフー等のインターネット検索システムとを包括する分類検索システムの一般設計論の必要性を試みる。特に、本論文では、上記分類検索システムの一般設計論の構築におけるグラフ理論適用の可能性を考えることを目的とする。

2. 特許分類の必要性と問題意識

ある発明が記載された特許出願は、特許庁に出願された後、審査官により、その発明の特許性の有無が審査され、特許性有りとは判断されれば特許権が付与され、逆に、特許性無しと判断されれば拒絶査定がされ特許権は付与されない。この審査における特許性の有無の判断は、出願に係る発明が法定の特許要件（新規性、進歩性、産業上利用性等）を具備しているか否かを判断することを内容とする。特に、法定の特許要件の中でも、「発明の進歩性」の有無を判断することが実態上、最も重要である。この「発明の進歩性」の具体的内容は、特許法第29条第2項の規定上、「特許出願前にその発明の属する技術の分野における通常の知識を有する者が前項各号に掲げる発明（例えば、出願前に公然知られた党の理由で新規性を有しない技術：「従来技術」や「先行技術」ともいう。著者が挿入）に基いて容易に発明をすることができた」か否かを判断することである。この規

定中「通常の知識を有する者」を「当業者」ともいう。即ち、審査官が特許出願に係る発明の進歩性を判断するとは、特許出願前の従来技術（主に、過去に出願・発行された特許公報に記載された技術）を調査し、一件の一番近い従来技術をピックアップし、出願に係る発明とその従来技術とを比較し、出願時点における発明の属する技術分野における当業者の立場に立って、「その従来技術に基いて容易に発明をすることができた」か否かを判断することである。

一方、現在、過去に出願・発行された特許公報（実用新案公報を含む。）は世界中で5,000万件以上あり、原則的に、一件の特許出願に係る発明の進歩性を正確に判断するためには、これら5,000万件以上の特許公報を全て調査しなければならない。しかし、一人の審査官がこれら全ての特許公報を調査することは現実的に無理である。そこで、以前から、特許公報に審査のための分類即ち特許分類が付与されており、特許出願に係る発明の進歩性を判断する場合、まず、所定の特許分類を指定してその指定された特許分類に該当する特許公報を調査し、その中から一件の一番近い従来技術をピックアップし、特許出願に係る発明とその従来技術とを比較し、発明の進歩性を判断している。従って、特許分類の善し悪しが発明の進歩性の判断の善し悪しに直結する。即ち、所定の特許分類を指定しても「漏れ」即ち「本来一番近い従来技術なのに、この従来技術が所定の特許分類内に含まれていない場合」が多くては、本来進歩性を有しない発明に特許権が付与され、審査の質が落ちることになる。逆に、所定の特許分類を指定しても「ノイズ」即ち「本来調査しなくてもいい従来技術なのに、この従来技術が所定の特許分類内に含まれている場合」が多くては、調査時間が長くなり、審査機関が長期化することになる。従って、審査の質の維持・向上及び審査期間の短縮化の両立が求められている今日、審査・調査のための特許分類を如何に設計すべきか、即ち、審査の質の維持・向上のためには「漏れ」の少ない又は全くない特許分類及び審査期間の短縮化のためには「ノイズ」の少ない又は全くない特許分類が理想的であるが、この理想的な特許分類を如何に設計すべきか、ということが大きな課題となる。

また、「特許分類の設計」といっても、「特許分類自体の構造」及び「特許分類の付与の仕方」という特許分類自体の客体的性質に関する側面のみならず、この特許分類を利用して調査・検索そして審査する審査官の認知過程に関する側面をも問題にする必要がある。即ち、上記した如く、発明の進歩性判断において、最終的に、一件の一番近い従来技術をピックアップし、特許出願に係る発明とその従来技術と

を比較し、「出願時点」における「出願に係る発明の属する技術分野」における「当業者」の立場に立って、発明の進歩性の有無を判断する必要がある。従って、発明の進歩性の判断時点において、審査官は「出願時点」における「出願に係る発明の属する技術分野」を正確に特定し、この技術分野の「出願時点」における「当業者」の立場に正確に立つことを要する。

もちろん、「出願に係る発明」と「過去に発行された特許公報に記載された従来技術」間において、付与された特許分類が同一であれば、概ね、技術分野が同一であるという推定は働くが、新たな技術技術が次々と開発・出願されている今日、たとえ付与された特許分類が同一であっても、「出願に係る発明」の属する技術分野が「過去に発行された特許公報に記載された従来技術」の属する技術分野と異なったり、又は大きく重複はするが微妙に異なる場合もある。従って、正確な進歩性判断のためには、出願に係る発明の技術理解——→特許分類の決定——→該当する特許分類における従来技術の調査——→一件の一番近い従来技術をピックアップ——→出願に係る発明と従来技術の比較——→発明の進歩性の有無の決定という一連の発明の進歩性判断過程において、特許分類の異同に拘泥することなく、審査官は、「出願時点」における「出願に係る発明の属する技術分野」を正確に特定することを要する。即ち、上記一連の発明の進歩性判断過程において、審査官が「出願時点」における「出願に係る発明の属する技術分野」を正確にかつできるだけ早く特定するためには、いかなる特許分類が妥当であるか、という観点が重要になってくる。従って、審査官の特許分類の習得・利用・開発のなかで、即ち、審査官の認知過程において、審査官は「出願時点」における「出願に係る発明の属する技術分野」をいかに正確に特定していくのか、という点に焦点を当てる必要があるとなる。

また、上述した如く、発明の進歩性の判断時点に、審査官は、「出願時点における当業者」の立場に立つことを要する。従って、上記一連の発明の進歩性判断過程において、審査官が「出願時点における当業者」の立場に立つためには、いかなる特許分類が妥当であるか、という観点も重要になってくる。従って、審査官の特許分類の習得・利用・開発のなかで、即ち、審査官の認知過程において、審査官が「出願時点における当業者」の立場をいかに正確に形成していくのか、という点に焦点を当てる必要があるとなる。

また、上述した如く、「漏れ」や「ノイズ」のない特許分類を設計し、「出願時点」における「出願に係る発明の属する技術分野」や「当業者」を正確に特

定し、最終的に、審査官は、従来技術に「基いて容易に発明をすることができた」か否かという客体的内容を判断することを要する。従って、発明の進歩性の判断時点において、審査官は、この客体的内容を判断するための沈着冷静な自分というものが即ち確かなアイデンティティを持つことを要する。心理的又は肉体的に疲労が重なっては上記客体的内容を正確に判断することはむずかしい。例えば、「異なる特許分類からさらに近い従来技術が出てくるかもしれない」とか「自分の判断に今ひとつ自信が持てない」という不安感を持った状態では正確な判断をすることがむずかしい。従って、審査官自身が自己の判断に自信をもつためには、いかなる特許分類が妥当か、審査官自身が自己成長を感じ具体的にキャリアアップできるためには、いかなる特許分類が妥当か、審査官自身が過去のノウハウを生かすことができるためには、いかなる特許分類が妥当か、審査官自身が過去のノウハウを生かし自己成長を感じ自己の判断に自信を持ち安定感のある自分を感じることができるといふためには、いかなる特許分類が妥当か、という観点も重要になってくる。従って、審査官の特許分類の習得・利用・開発のなかで、即ち、審査官の認知過程において、審査官自身が自己のアイデンティティというものをいかに有意に確立していくのか、という点に焦点を当てる必要がある。

現在、日本においては、主に、三種類の特許分類IPC・FI・Fターム（詳細は後述）が使用されている。このうち、Fタームは、昭和59年からのペーパーレス計画の実現に伴い、開発・設計されたものである。このFタームは、審査官等の長年の経験から開発されたもので、実態上、実態上、発明の進歩性の判断過程において、非常に効果的で、現在の特許出願の審査期間の短縮化及び審査の質の維持・向上に大きく役立っている。実態上、Fタームは特許分類の設計・開発における大きな成功事例であるが、なぜ、Fタームが特許分類の設計・開発における大きな成功事例であるのか、なぜ、Fタームが発明の進歩性の判断過程において非常に効果的であるのか、という思想体系は現在、存在していない。また、特許調査・審査以外の調査目的が明確でかつ特に高い再現率・適合率が要求される調査・検索を行うデータマイニング即ちアクティブマイニングにおける分類の必要性・有効性等に関する学問体系が存在していないのが実情である。インターネット技術の発達に伴い、これから進めていく電子政府の実現や多くの分野におけるペーパーレス化の流れのなかで、一步先んじて、ペーパーレス化及びこれに伴うFターム検索システムを実現した開発成功事例を体系化・一般化し、他の分野に応用・移植していく

ことは有益なことである。

3. 細分類と再分類の図示（ともに、読み「サイブ
ンルイ」は同じ）

3. 1. 細分類（アナログ分類）のイメージ

IPCは、全技術分野を大きく次の8セクションに分けている。

Aセクション：生活必需品

Bセクション：処理操作；運輸

Cセクション：化学；冶金

Dセクション：繊維；紙

Eセクション：固定構造物

Fセクション：機械工学；照明；加熱；武器；爆破

Gセクション：物理学

Hセクション：電気

上記セクションから派生して、階層構造になっており、さらなる下位分類として、順にサブセクション→クラス→サブクラス→メイングループ→サブグループと細分類されている。

例えば、

G：物理学

G06：計算または計数

G06F：電氣的デジタルデータ処理

G06F16/00：情報検索；そのためのデータベース構造；そのためのファイルシステム構造

G06F16/80：・半構造化データのもの、例. SGML, XMLまたはHTMLなどのマークアップ言語構造化データ

G06F16/81：・・・インデキシング、例. XMLタグ；そのためのデータ構造；ストレージ構造

3. 2. IPCの基本的付与方法（概要）

(1) 発明は、全体として分類される。

発明はその構成部分を個々に分解して分類することなく、全体として分類される。

(2) 機能指向箇所と応用箇所

発明の技術主題は2つのタイプに区分される。

① 機能指向的なもの

固有の性質又は機能に特徴を有するもの、ある特定の使用分野に限られないもの、使用分野の記載を無視しても技術的に影響のないもの

② 応用指向的なもの

特別の使用又は目的に特に適合したもの、特殊な使用又は応用、より大きな系への組み込み

上記①、②の分類箇所は各々、機能指向箇所、応用箇所と呼ばれる。発明の本質的な技術主題を特定した上で適切な分類箇所に分類する。技術主題の本質的な技術的特徴が応用がない場合は機能指向箇所に分類し、上記①と②の両方に関係する場合は機能指向箇所と応用箇所の両方を分類する。

実態上、一件の特許出願に付与されるIPCの数は1個から数個程度である。

付与の具体例 B 2 1 D 2 2 / 1 0

3. 3. FIの具体的内容(図1参照)

3. 3. 1. FIの具体的構成

FI (File Index) とは、上記IPCのみでは不十分な分野において、原則として、上記IPCの最小の分類単位であるサブグループをさらに詳細に細分類したものである。

3. 3. 2. FIの基本的付与方法

原則として、上記IPCの基本的付与方法と同様である。

付与の具体例 B 2 1 D 2 2 / 1 0 A

3. 3. 3. Fタームの具体的内容(図1及び図3参照)

3. 3. 3. 1 Fタームの具体的構成

Fターム (File Forming Term) は、機械検索用に開発された特許分類である。上記IPCは単一の技術的観点を中心に展開されており、技術の分けも粗い。このため、IPCのみでは、特に最近の発展した技術分野においては数百件から数千件の従来技術を調査することを要する。従って、Fタームは種々の技術的観点(目的、用途、構造、材料、製法、処理操作方法、制御手段等)からIPCを所定技術分野(テーマ)毎に再区分又は細区分したもので、従来技術を50から70件程度に絞り込むことを目指している。次の三種類のFターム方式がある。

(1) FS方式——技術の多様化した分野でのサーチ等、FIのみでは効率的

でない場合に、各FI毎にそれぞれサーチに必要なタームを細展開する方式

(2) FM方式——技術の複合化した分野での横断的サーチ等、多くのFIを使う必要がある場合に、FIで規定された技術範囲にわたりFIとは異なる複数の観点で再区分したタームを展開する方式。

(3) ハイブリッド方式——上記(1)及び(2)の複合型

3. 3. 3. 2 Fタームの基本的付与方法

「種々の「観点」の分類タームをなるべく多く付与し、サーチの都度データベースを利用して電算機中に必要なファイルを作る」という考えに基づいてFタームの開発・付与が行われている。FIで区別したテーマの数は全技術分野で約3,000前後で、その内の大部分は上記(2)のFM方式である。尚、FIのみで十分な検索ができる場合は、Fタームは作成されない。

付与の具体例 5 B 0 0 3 A A 0 3

(「5 B 0 0 3」はテーマコード)

4. 細分類(アナログ分類)に関するグラフ理論と再分類(デジタル分類)に関するグラフ理論との関係

4. 1. 複数のグラフ(アナロググラフとデジタルグラフ)との重複・適用関係

4. 1. 1. 定義

アナログ分類に対するグラフ理論適用⇒アナロググラフ

デジタル分類に対するグラフ理論適用⇒デジタルグラフ

まず、アナロググラフを適用

次に、デジタルグラフを適用

4. 1. 2 グラフ理論を適用する際の基本的枠組(体系分類における細分類と再分類)

4. 1. 2. 1 分類項目の在り方に関して

- ・細分類⇒既存分類項目をさらに細かく分類する。
- ・再分類⇒複数の既存分類項目をひとまとめにして、このひとまとめにした複数の分類項目に含まれる文献のみを対象にして異なる観点から新たな分類項目を設定して分類し直す。

4. 1. 2. 2 アナログ分類とデジタル分類

- ・アナログ分類⇒IPC(国際特許分類)、FI(FILE INDEX)(IPCを細分類したもの)⇒演繹的な体系性を重視⇒最上位概念から演繹的に、より下位に向けて段階的に細分類していく。

- ・デジタル分類⇒Fターム(FILE FORMING TERM)⇒上記アナログ分類において最上位概念から演繹的に、より下位に向けて段階的に細かく分類した所定の複数の細分類項目をひとまとめにして(このひとまとまりを「テーマコード」という。現在、約2千数百個のテーマコードが存在する。)、ひとつのテーマコード内において異なる観点から新たな分類項目を設定して分類し直す(再分類)。
- ・従って、順序としては、まず、アナログ分類において細分類、次に、デジタル分類において再分類(細分類⇒再分類)。

4. 1. 2. 3. 各分類の付与方法に関して

- ・アナログ分類⇒1個又は2, 3個程度のアナログ分類項目を付与する。

- ・デジタル分類⇒できるだけ多くのデジタル分類項目を付与する。

4. 1. 2. 4. アナログ分類に関するグラフ理論とデジタル分類に関するグラフ理論との関係

- ・アナログ分類を前提としたデジタル分類しか使い物にならない。従って、デジタル分類を前提としたアナログ分類は有り得ない(まず間違いなく)。また、アナログ分類を前提としないデジタル分類だけでは、

体系分類としては使い物にならない。

4. 2. ナレッジグラフ (様々な情報を、意味を表す関係性でつなぎあわせたグラフ型の構造化知識)

4. 2. 1. 研究の方向性に関する重要概念

- ・分類検索における「分類」⇒文章群の意味概念
- ・テキスト検索における「テキスト」⇒言語・単語の意味概念

従って、両者は、意味概念であることは共通するが、概念の安定性・揺れの大小が全く異なる。

もちろん、両者は、意味概念の対象が明確に異なる。即ち、分類検索における「分類」は文章群を対象にしているのに対し、テキスト検索テキストを対象にしている。付与方法も明確に異なるように思う。即ち、分類検索における「分類」は人手によるアナログで分類するのに対し、テキスト検索における「テキスト」はコンピュータによるデジタルで分類する(又は積極的に分類しない)。ただ、テキスト検索における対象文献の文章を入力する際、後のテキスト検索のために統一語を入力する場合があるが、この点は分類検索における事前付与と共通する作業要素を含んでいる。

4. 2. 2. 前提となる基本概念

4. 2. 2. 1. 検索の種類

(1) キーワード(テキスト)検索(テキストデータに対して、機械的にキーワード検索を行う。例えば、ヤフーやグーグル等のインターネット検索)

① 長所⇒キーワードは分かりやすいため、初心者等でも容易に検索しうる。事前に体系分類の分類タームが付与される必要がないので、体系分類を使用した業務と比較するとキーワード(テキスト)検索を使用した業務には事前に分類付与特有の時間とコストが必要とされない。

② 短所⇒類義語等への読み替えはなされないため、分類検索と比較すると、サーチ漏れの可能性が大きくなる。関連性の小さい文献でもキーワードが存在する場合があるのでノイズが多くなる。

(2) 分類検索(体系分類の分類タームを使用した検索、例えば特許分類検索)

① 長所⇒事前に人間が文献を読み内容を理解して分類タームを付与するので、類義語や誤記が存在しても適切な分類タームが付与されることにより、サーチ漏れやノイズが少なくなる。

② 短所⇒体系分類における体系性や分類構造を理解する必要があるため、初心者にとっては使用しづらい。事前に分類タームを付与する必要があるため、体系分類を使用した業務に時間とコストが必要とされる。

4. 2. 3. 体系分類とナレッジグラフとの関係性?

4. 2. 4. キーワードとナレッジグラフとの関係性?

4. 2. 5. 2つ以上のナレッジグラフの重複適用の可能性?

4. 2. 5. 1. アナロググラフとデジタルグラフの順序関係?

4. 2. 5. 2. アナロググラフとデジタルグラフの重複関係?

図1 IPC, FI, Fタームの関係 (「ペーパーレス計画及びFターム検索システムについて」(財)工業所有権協力センター発行、平成10年発行の10頁の図2を転載)



