

分類検索システムとグラフ理論との関係

体系分類の可能性・限界 VS 自然言語処理の限界・可能性

ノウハウとアイデンティティの維持発展を追い求めて

Relation of Classification-Retrieval-System and Graph-Theory

Systematic Classification VS Natural Language Processing

In pursuit of Importance of Know-how and Identity

上田 育弘

Ikuhiro Ueda

ベストフレンド知的財産研究所

Best Friend Intellectual Property Laboratory

(要旨) Fターム検索システムは、分類検索の極致にあるシステムで、キーワード検索（自然言語処理）を徹底的に排除している。一方、グーグル・ヤフー等のキーワード検索システムは、キーワード検索（自然言語処理）の極致にあるシステムで、分類検索を徹底的に排除している。特許分類 IPC・FI・Fターム検索システムにおける検索者のノウハウ維持とアイデンティティ強化の成功要因・成功事例は、今後の人間中心の情報システムにおける人間・検索者のノウハウ維持とアイデンティティ強化のひとつの具体例であるともいえ、今後は、Fターム検索システムの成功要因の理論化・一般化により、人間中心の情報システム構築の大きな知見を得ることが可能である。

1. 問題提起（はじめに）（詳細は参考文献をご一読下さい）

(1) 発明の進歩性判断のための従来技術調査の特質

迅速かつ的確な審査・調査が求められている。

① 迅速な審査・調査のためにノイズの最小化の必要性：適合率=1が理想

② 的確な審査・調査のためにモレの最小化の必要性：再現率=1が理想

A：全ヒット件数、B：本来欲しい情報の件数、C：ヒットした本来欲しい情報の件数

★適合率=C/A=1が理想、・再現率=C/B=1が理想

◎A-C=ノイズ件数・B-C=モレ件数

(2) 特許分類 IPC・FI と Fターム との比較

IPC：International Patent Classification（国際特許分類）

FI：File Index

Fターム：File Forming Term

	IPC・FI	Fターム
目的	新規性・進歩性判断に必要な従来技術調査を行う。	新規性・進歩性判断に必要な従来技術を約50件に絞る。
分類構造	発明に関する全技術分野を段階的に細分化した技術分類	機械検索用に開発された特許分類で、種々の技術観点（目的、用途、構造、材料、製法、処理操作方法、制御手段等）から細分化又は再分類化
分類手法	細分類（細かく分ける）	細分類（細かく分ける）又は再分類（分類し直す）

テーマ（枠概念） 設定の有無	テーマ設定無し	IPC・FIのなかの所定の特許分類のみを対象にし、その対象範囲をテーマとして設定、現在約3,000個のテーマが存在
分類付与方法	該当分類を付与し、1,2個から多くても数個	できるだけ多くの分類を付与する、10個以上も珍しくない
分類の体系性	最上位の8分野（生活必需品・処理操作運輸・化学冶金・繊維紙・固定構造物・機械工学等・物理学・電気）から順次、ひたすら細分類（細かく分ける）	機械検索用に開発された特許分類で、種々の技術観点（目的、用途、構造、材料、製法、処理操作方法、制御手段等）から細分化又は再分類化

（3）検索の種類

①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IPC・FI・Fターム検索

- ・（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。

- ・（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。

② キーワード検索（自然言語処理）⇒キーワードを使用した検索、例えば、グーグル・ヤフー等のキーワード検索システム

- ・（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。

- ・（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。

★Fターム検索システム⇒分類検索の極致にあるシステムで、キーワード検索（自然言語処理）を徹底的に排除している。

★グーグル・ヤフー等のキーワード検索システム⇒キーワード検索（自然言語処理）の極致にあるシステムで、分類検索を徹底的に排除している。

◎Fターム検索システムの成功要因

①テーマ設定したこと②多観点に展開したこと③審査官のノウハウを温存し審査官個人のアイデンティティの確立することに有用であること、即ち、テーマ設定しているため、このテーマに関する専門技術を有する審査官が自己の有する過去から現在まで蓄積してきた体系性あるノウハウを進歩性判断のための従来技術調査に直接活用しつつ、多観点に展開してできるだけ多くのFタームコードを付与しているため、流動する技術需要に柔軟に対応することができる。

2. ナレッジグラフ及びグラフ理論と分類検索システムとの関係性に関する試論

（1）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

(2) グラフ理論⇒「点と線」で構成された図形をグラフというが、グラフの構造を調べたり、その応用を研究する学問がグラフ理論である。この学問は専門家以外の素人でも理解できる問題が多いことと、応用領域が広くて多様である点に特色がある。グラフ理論は、数学の一分野であるが、トポロジーと組み合わせ理論の境界あたりに位置する学問である。トポロジーが純粋数学の立場からグラフを捉えているのに対し、組み合わせ理論は応用数学の立場に立って、グラフを学問や技術の他の世界に役に立てることを目指している。

(3) 特許分類IPC・FIと新特許分類Fタームにおける各体系性

①特許分類IPC・FIの体系性⇒「最上位の8分野（生活必需品・処理操作運輸・化学冶金・繊維紙・固定構造物・機械工学等・物理学・電気）から順次、ひたすら細分類（細かく分ける）」⇒従来から存在する既存の体系性を維持しているので、従来と同様に、組織分類として利用できる（最終頁の組織担当表参照：日本特許庁における審査第一部～審査第四部の分類担当は特許分類IPC・FIの体系性を根拠としている）。

② 新特許分類Fタームの体系性⇒「機械検索用に開発された特許分類で、種々の技術観点（目的、用途、構造、材料、製法、処理操作方法、制御手段等）から細分化又は再分類化」⇒流動性が大きく、組織分類としては利用しにくい。

③ 従って、Fターム検索システムは、特許分類IPC・FIの体系性を前提としつつ、新たに約3000のテーマ毎に新特許分類Fタームの体系性を新設した構造をしている。

(3) 上記二種類の体系性のグラフ表現たる接続行列（参考文献：本間龍雄著「グラフ理論入門・点と線の数学」30頁乃至33頁参照）

- ・体系分類は、正方行列になる。
- ・体系分類は、対称行列になる（±逆になる）。
- ・体系分類は、0、+1、-1の3つのみ。
- ・体系分類は、一点から出発して枝分かれしていき、必ず、端点で終わる。

・上記した体系分類に関する4つの内容は、特許分類IPC・FIと新特許分類Fタームとの間で同一であるが、付与方法が全く異なる。結局、特許分類IPC・FIで大枠（テーマ）を決定し、この大枠（テーマ）のなかで異なる体系性を有する新特許分類Fタームを付与し再分類することにより、キーワード検索（自然言語処理）の短所たるサーチ漏れ及びノイズを極力少なくしつつ、検索者のノウハウやアイデンティティを保持しようとしたのが、Fターム検索システムといえる。

従って、今後は、キーワード検索（自然言語処理）の長所たる「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。」を如何に生かしていくかが、Fターム検索システムにおける大きな課題となる。

3. 人間中心の情報システムにおける人間・検索者のノウハウとアイデンティティとの関係

(1) 情報システム学会メールマガジン 2021. 11. 30 No. 16-08 連載 情報システムの本質に迫る

第174回情報システムとは何かー情報システム学会の考え方（吉賀正憲著）から抜粋

	コンピュータ	組織・社会
組織・社会そのものの仕組み		人間中心の情報システム 基礎情報学、リベラルアーツ、哲学、心理学、言語学、文化人類学、社会学、

		経済学、統計学、経済学、I E、人間工学、生産工学、・・・ビジネスモデル、業務モデル・・・などにより説明されてきている
情報処理の仕組み	コンピュータによる情報処理の仕組み	組織・社会における情報処理の仕組み

(2) 上記情報システム学会の考え方における「情報システム」の定義に対する本質的疑問
 ・「コンピュータ」「組織・社会」の二項目に焦点が当てられているが、これら二項目間の具体的な相関関係に焦点が当てられていない。

・「コンピュータ」の項目に焦点が当てられているが、「コンピュータ」を考えた人間・個人及び「コンピュータ」を利用する人間・個人に焦点が当てられていない。

・「組織・社会」の項目に焦点が当てられているが、「組織・社会」を構成している人間・個人に焦点が当てられていない。

・「組織・社会そのものの仕組み」「情報処理の仕組み」「コンピュータ」という客観面に焦点が当てられているが、「組織・社会」を構成している人間・個人並びに「コンピュータ」を考えた人間・個人及び「コンピュータ」を利用する人間・個人そして「情報処理」をコンピュータに行わせる人間・個人という各人間・個人の主観面（目的・やる気・意欲・ノウハウ・アイデンティティ等）及び各人間・個人の「業務」自体に焦点が当てられていない。

・「コンピュータ」出現以降の「コンピュータ」「組織・社会」の二項目を中心に概念が構成されており、「コンピュータ」出現以前の各人間・個人の主観面（目的・やる気・意欲・ノウハウ・アイデンティティ等）及び各人間・個人の「業務」自体に焦点が当てられていない。

・「組織・社会そのものの仕組み」「情報処理の仕組み」「コンピュータ」という客観面に焦点が当てられているが、自由主義・社会主義・個人主義・共産主義等の人間・個人の考え方・思想自体に焦点が当てられていない。

・技術史・思想史・産業革命史（第1次（18世紀中盤）・第2次・第3次・第4次・第5次～）と「組織・社会そのものの仕組み」「情報処理の仕組み」との間の相関関係等の歴史的考察が欠落している。

・代表的なユーザたる会社組織内の経営者・開発部門・現業部門の三者間の関係やユーズとベンダとの関係等の具体的な関係論が欠落している。

(3) 特許分類 IPC・FI・F ターム検索システムにおける検索者のノウハウ維持とアイデンティティ強化の成功要因・成功事例は、今後の人間中心の情報システムにおける人間・検索者のノウハウ維持とアイデンティティ強化のひとつの具体例であるともいえ、今後は、F ターム検索システムの成功要因の理論化・一般化により、人間中心の情報システム構築の大きな知見を得ることが可能である。

4. 参考文献

・上田育弘著「審査官の認知過程に着目した「特許分類IPC・FI・Fタームを活用したアクティブマイニング」人工知能学会研究会資料SIG-FAI/KBS-J-03(11/12)～特許分類IPC・FI・Fタームを活用した審査官の進歩性判断過程に認知過程における一般的モデルである多重記憶モデルを適用し、Fターム検索システムの成功要因として①テーマ設定したこと②多観点に展開したこと③審査官のノウハウを温存し審査官個人のアイデンティティの確立することに有用であることの3点を挙げている。～

・本間龍雄著「グラフ理論入門・点と線の数学」、昭和59年7月15日第8刷・株式会社講談社発行

★日本特許庁における審査第一部～審査第四部の分類担当

審査第一部	距離・電気測定、物理測定、表示装置、医学診断、画像診断、光学装置、光学要素・EL素子、光制御、印刷・プリンター、電子ゲーム・学習、アミューズメントマシン、アミューズメント一般、住宅設備
審査第二部	電動機制御、流体制御、タービン・機関部品、駆動制御、車両制御、制動・機械要素、加工機械、ロボティクス、端子部品、包袋・容器、福祉・サービス機器、照明、冷却機器、治療機器
審査第三部	セラミックス、蒸着・単結晶成長、樹脂加工、電気化学、電池、バイオ医薬、医薬品製剤、食品・微生物、蛋白質工学、分離処理、複素環化合物、農薬・染料、塗料・接着剤、高分子組成物、繊維・積層
審査第四部	金融・決済、ヘルスケア情報システム、検索・記憶管理、イメージ処理、情報セキュリティ、伝送回路、移動体通信システム、送配電、データネットワーク、テレビジョン、静止画、音響システム、デバイス実装、デバイスプロセス