

NRI サイバーパテント

特許の 質を見る

進化するTRUE TELLER パテントポートフォリオ

NRIサイバーパテントの提供する「TRUE TELLERパテントポートフォリオ(TTL-PP)」は、野村総合研究所(NRI)グループが、テキストマイニングの先進活用分野である企業のCS部門^{注1}・マーケティング部門において180社以上の導入実績を有する「TRUE TELLER」の技術・ノウハウを、公報の解析に応用した“特許ポートフォリオ分析専用アプリケーション”である。NRIグループのシステム構築力と、知財分野における調査・コンサルティングの蓄積・ノウハウの結晶であるTTL-PPは、2004年10月の知財分野への展開開始以降、すでに40社近くの有力企業で活用されているが、分析手法・機能の進化は止まらない。

注1 CS部門：CS（Customer Satisfaction：顧客満足）の推進を行う企業内の組織。顧客が取引継続を続ける場合、商品やサービスに対する顧客満足度が重要になる。CS度をアップさせるにはコールセンターや顧客対応窓口の設置によって顧客のクレームに対応する一方、常に顧客意識を調査または情報を提供してコミュニケーションを取り続けることが必要。

注2 サーモグラフ：気温や体温の違いの分布状態を色で区分けして表現する方法。たとえば高温を赤、低温を青として、二色間の色階調で細かく表現する。この手法の特許文献や単語の密度に応用したのがNRIサイバーパテントのサーモグラフ機能である。

知財ポートフォリオ 可視化の新技法

5月末に次期バージョンアップも

NRIサイバーパテントはTTL-PPのポートフォリオ分析機能に、このほど新たに発明の「質」を反映するための機能を追加する。

TTL-PPのもっとも特徴的な分析手法の一つである「サーモグラフ^{注2}」分析は、はじめに母集団の公報群に類出する単語間の同時使用頻度をもとに、単語マップを自動作成する。この単語マップを共通の下敷きとした上に、出願人毎の出願領域の違いや、時系列での推移を、文献数の密度で表現するのがサーモグラフである。公報の文中に含まれる言葉を見ることで、これまでの特許分類などをキーにしたパテントマップでは難しかった、発明の内容まで踏み込んだ、特許ポートフォリオの可視化手法として好評を得てきた。

新機能は、1件1件の特許・出願に、被引用文献数、閲覧請求回数などをパラメータ（分析する際の変数項目）として重み付けすること

で、基本特許としての性格、ライバルからの注目度といった発明の「質」を加味した分析を可能とした（図表1）。これにより、企業×技術領域毎の強さ・弱さを、よりの確に把握することができる。パラメータについては、こうしたNRIグループからの提案のほかに、自社で独自の評価点（スコア）・ロジックなどを保有している場合にはこれを反映することもできる。被引用文献数、閲覧請求回数などのデータは、インターネット特許情報サービス「NRIサイバーパテントデスク」から簡単に取得可能だ。

また、サーモグラフ分析は、単に競合企業間のポートフォリオの違いを横並びに比較するだけでなく、密度分布の「差分」表示によりベンチマークになる企業に対して自社が強い領域を浮き彫りにしたり、または「合算」表示により他社と提携した場合のポートフォリオの変化をシミュレーションするといった使い方もできる。

NRIサイバーパテントでは今後も機能追加をしていく考えで、「5月末には次期バージョンアップを予定しており、テキストマイニングと引用分析（サイテーションツリーなど）の組

図表1 リチウムイオン電池に関する分析例

合せより、分析機能のさらなる高度化を進めていく予定」(企画営業部・中居隆氏)としている。

単語・係り受けから、文献をグルーピング

図表2は、公報中に頻出する単語および係り受け(主語-述語、修飾語-被修飾語の関係)から、特許を、その内容によってグルーピングし、グループ間のポートフォリオを分析している。たとえば、出願人×課題グループの距離を見ることで、自社と他社の技術戦略の違いを浮き彫りにしたり、出願人別の課題×用途のバブルチャートから、他社の取り組みの少ない「空き地」を見つけることにも役立つ。

こうした分析は、企業によっては、これまでも大量の人手と時間を介して行われてきた。

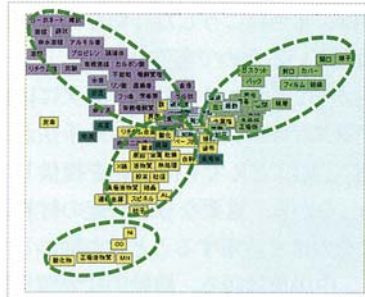
たとえば、目の前に500件の公報の印刷物があり、その特許が扱う課題によって分類していくとしよう。担当者は、まずは50件ほど、各特許を特徴付ける技術用語のキーワードにマーカーを付けながら読み進め、そこで、整理の枠組みを仮決めし、文献を机上で仕分けしていただく。ところが、さらに200件まで分類を進めると、新たなカテゴリーの存在、あるいは、何度も出てきたが重要と思っていなかったキーワードの存在に気づき、一からやり直しということになる。技術の多様化とクロスオーバーが進む中で、新しい研究開発を始める際に目を通すべき特許が数千件を超えれば、こうした腕力に頼ったやり方は事実上不可能となっている。

TTL-PPによる文献分類は、まさに「大量の特許公報を読み込んで分類・分析する」プロセスを、もとの文献に紐付いた頻出単語・係り受け(主-述、修-被修)の選別により、極めて効率的に行うものである。TTL-PPが保有する同義語辞書、類義語の抽出機能、単語の部分一致指定機能が、表現の揺れの吸収をサポートする。

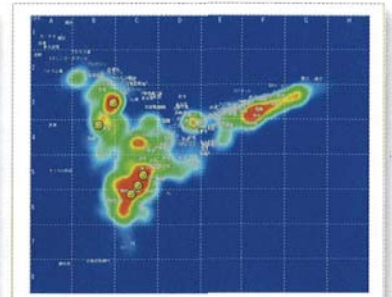
問題意識・仮説に基づいて詳細分析

「誤解を恐れずに言えば、TTL-PPによる分析は、データを取り込むと自動的に結果を出してくれるといった類のものではない」と中居氏は説明する。たとえば、文中に含まれる言葉・係り受けから、特許文献をグルーピングするには、その技術領域に一定の知識・見識のあ

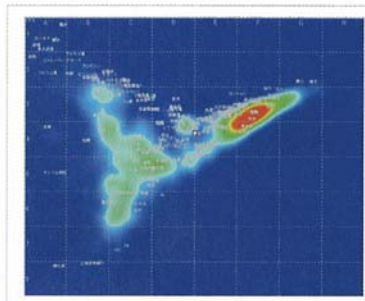
単語マップ(下敷き)
(単語に付与した「属性」によるグルーピング)



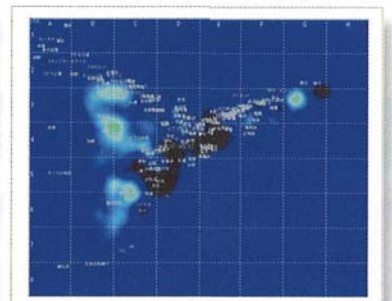
A社 被引用数で重み付け
(プロットは被引用数≥10の文献)



A社 単純件数



A社-B社
(緑~水色: A社が出願多、
紺~黒: B社が出願多の領域)



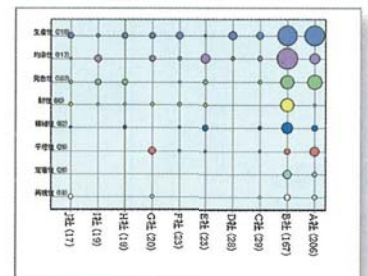
図表2 インクジェット技術に関する分析例

課題グループの設定

課題グループ	文中に含まれる単語・係り受け
生産性 (218)	生産性、量産、経済性、コスト価格、安価、安い など
均染性 (117)	均染、むら、滲み、混色、画素-欠陥、均一分布、均一付与、色-抜け など
耐性 (90)	耐光、耐水、耐熱、耐薬品、耐候、耐湿、耐衝撃、保存、気温、湿度、湿気、保存する、耐える など
精細性 (62)	鮮明、精細、解像度、細かい、粗い など
平坦性 (29)	平坦、平滑、平ら、均一-厚み、均一-厚さ など
定着性 (28)	密着、定着、染着 など
再現性 (18)	再現、色調、色-純度 など
発色性 (07)	発色、発光、光沢、濃色、コントラスト、鮮やか、色-濃い など

頻出単語・係り受けからユーザの視点で特許をグルーピング

出願人×課題グループ#1



出願人×課題グループ#2



る知財担当者あるいは技術者・研究者の目が不可欠である。また、分類の切り口についても、特許の扱う課題、機能、用途、構成要素、製法など、企業が抱える問題意識、仮説に基づき、さまざまな視点から設定できる。出願審

特許の「質」を見る
進化する TRUE TELLER パテントポートフォリオ

査のために付与された「特許分析」による分析が、そのまま企業の意味決定の材料としてなじみにくいのは、まさにこうした「切り口」、「視点」が欠如しているためである。

同じ理由から、「テキストマイニングには自動分類（クラスタリング）と呼ばれる手法があり、TRUE TELLERでも同機能を提供している。しかしながら、重要な意思決定の材料となる分析にそのまま活用することはお勧めしていない」と、中居氏は語る。機械的に分類を行うロジックが部分的にブラックボックス化されるため、ユーザーにとって明確な説明・検証が難しくなるためだ（マーケティング分野で活用される場合は、主に、担当者レベルでの母集団

の確認・把握にのみ使用されている）。

TTL-PPの採用・推奨する分析ロジック・プロセスは、極めてオープンなものである。導入企業を見ると、特許情報の扱いに慣れた知財部門での活用が比較的多く、経営層、あるいは、現場の研究開発部門に対するアウトプットの客観性・網羅性・納得性を重視しているようだ。

**特許調査業務にも
テキストマイニング技術を活用**

目を通せる文献数には限界がある

NRI サイバーパテントが提供するテキストマイニング技術は、分析ツール＝「パテントマップツール」としてだけでなく、特許調査（検索）業務にも活用されている。

研究開発テーマの探索時や、新規テーマ候補に対する先行技術調査においては、特許検索のための条件の設定は手探りにならざるを得ない。人が目を通せる文献数には限界があるため、検索担当者は「500件以内に絞り込むために、検索式をどう設定するか」という本末転倒の思考をせざるを得ない。

TTL-PPでは、数万件～十万件以上の特許データ（要約、クレーム、全文等）を取り込み解析できるため、特許分類等によって設定した幅広い母集団に対して、文中に含まれる単語・係り受けから、本来、目を通すべき特許群を絞り込んでいくことができる（図表3）。一般に、検索式の設定において、「NOT式」は、本来目を通すべき文献を漏らす要因として敬遠されるが、TTL-PPではNOTではずした文献群について簡単に原文を参照でき、かつはずした集団の中に含まれる頻出単語を簡単に確認し、再び母集団に復活させるといったことが自由に行える。これまで大量の時間と労力を費やした

図表3 文中に含まれる単語・係り受けから母集団を絞り込み

The screenshot shows the TRUE TELLER software interface. The top window displays search results with columns for No., 単語 (Word), 品詞 (Part of speech), 頻度 (Frequency), and 割合 (Ratio). The bottom window shows a list of terms with checkboxes for selection and a search button.

No.	単語	品詞	頻度	割合
1	インク	名詞	2316	69.64
2	ジェット	名詞	1539	47.25
3	形成する	動詞	1520	46.19
4	提供する	動詞	728	22.25
5	提供	名詞	1154	35.20
6	飛ぶ	動詞	551	16.91
7	有する	動詞	564	17.48
8	製造方法	名詞	493	14.96
9	装置	名詞	601	18.47
10	パター	名詞	578	17.58
11	表面	名詞	441	13.84
12	印刷	動詞	456	14.15
13	材料	名詞	441	13.49
14	工程	名詞	520	15.95
15	含む	動詞	310	9.27
16	含む	動詞	310	9.27

「ブランド価値評価」情報提供サービスを開始

NRI サイバーパテントは、企業の財務情報と、知財等の無形資産を組み合わせた「企業ブランド価値評価」をインターネット上で閲覧できるサービスをこのほど開始した（<http://www.nri.co.jp/news/2006/060302.html>）。

経済産業省の企業法制研究会（ブランド価値評価研究会）で委員長を勤めた早稲田大学の広瀬義州教授と野村総合研究所が共同で実施している、「企業の知的財産評価プロジェクト」の一環として、企業のブランド価値評価データの算出をしている。このブランド価値評価データと特許情報を連動させたもの。同業他社・業界が保有する知的財産を、当該企業の財務状況、ブランド価値と重ねて分析することが可能となる。なお、利用価格は1件につき50円。

図表4 アラート分析機能

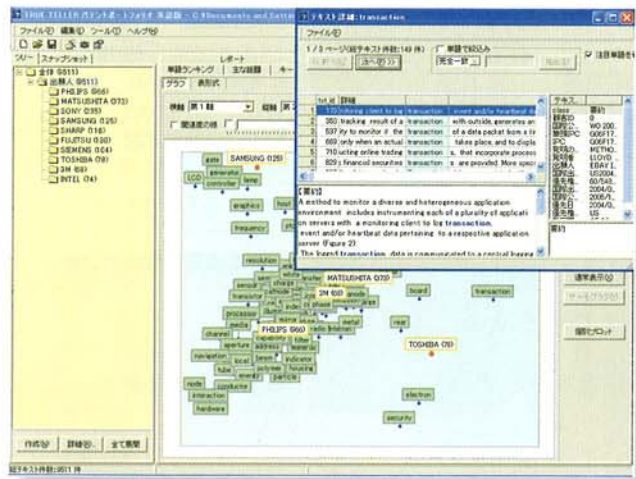
がら、「漏れ」のリスクを軽減しにくかった新規分野における特許調査にて、調査精度の向上と、コスト削減という視点からも効果が期待できるわけだ。

TTL-PPでは、CSV形式であれば、どんなデータでも取り込むことができるため、自社のデータベース、使用中のプロバイダなど、既存の資産・蓄積を活用することができる。目的に応じて、公報の要約・請求項・明細書全文を対象に分析することができる。また特許だけでなく、実用新案や技術論文データにも対応する。

絞り込んだ文献群や、すべてのアウトプットから、簡単に元データを参照できるほか、外部サービスへのハイパーリンクにより、図表付きの全文明細、経過情報をたどり、詳細に中身を確認でき、マクロからミクロへ、またはミクロからマクロへと柔軟な調査・分析対応ができる。



図表5 英語版の画面



継続的な調査業務への適用も

たとえば、研究開発を開始し、一定の成果を得て量産試行を開始するまでの間、企業ではいわゆる「SDI^{注3}検索」などを活用し、新着情報に目を凝らす。「SDI検索」は、指定した検索条件に対して新着情報を知らせてくれるツールとして有効だが、出願人が思い掛けない表現をしたり、まったく未知の技術用語が登場した場合にはお手上げとなる。

この不便さを解消する面白い機能がTTL-PPの「アラート分析」(図表4)だ。特許分類等により幅広く抽出した母集団に対して、全体あるいは出願人ごとに急に使用頻度が跳ね上がったキーワードを発見することができる。また、研究開発の開始段階において、課題・機能・用途等の切り口で設定したグループに対して、新着文献を自動的に振り分け、ここから漏れた文献群から、キーワードを抽出するといった使い方もできる。常に新しい用語をウォッチしていけるわけだ。

国際化対応で英語版も投入

知財競争が国際化する中で、むしろ海外特許の傾向・動向を把握したいという声が高まっているが、TTL-PPではこのほど、英語の解析エンジンを搭載した「英語版」の提供も開始している。

英語版は、日本語版とほぼ同様の分析機能

を提供しており、日本特許に加えて、米国・PCT等の文献を共通の手法・切り口から分析できる(図表5)。日本語版と同様に、CSV形式のデータであれば、データソースによらず解析が可能だ。

昨今、企業において業種・規模を問わず、知財部門への期待・要請が高度化し、大量の特許情報のハンドリングを余儀なくされていることが、テキストマイニングツールの導入が急加速している所以となっているようだ。インターネットによる特許情報サービスは1997年の登場以降、企業にとって当たり前の道具として急速に浸透していった。テキストマイニングもまた、特許分析・調査の不可欠な道具として浸透段階に入ったといえる。

注3 SDI : Selective Dissemination of Informationの略で、「情報の選択的配信手法」のこと。特許の世界では新しい特許公報やさまざまな情報を自動的に検索してきたり、配信を受けたりするためのツールやサービスを指している。これらは事前にニーズを設定した検索式に基づいて運用される。