特許情報サービスにおける画像検索

株式会社リコー　研究開発本部　先端技術研究センター　　伊東 秀夫

1. はじめに

リコーは長年にわたりテキストや画像を対象とする情報検索の研究開発を進めて参りました。その成果がリコーテクノシステムズ株式会社による特許情報サービス RIPWAY (tm) に活かされています。この商用サービスは、日米の特許公報を対象とした概念検索、言語横断検索、画像検索といった先端技術により特許調査の効率化にお役立ちすることを狙いとして2004年に開始されました。

本稿では、同サービスにおいて実用化した公報図面の検索機能（RIPWAYイメージサーチ）を通し、画像検索とは何か、その効果的な利用方法、検索の仕組み等についてご説明させていただきます。またこの技術を発展させることで実現した図形商標の検索機能についてもご紹介いたします。

2. 図面で図面を検索する

RIPWAYイメージサーチは、言わば画像の概念検索です。すなわち、データベースに登録された大量の公報図面群の中から、検索キーとして指定した図面に類似するものを検索し、類似度が大きい順に並べて検索結果として提示します。通常の概念検索との違いは、テキストではなく、画像の形状特徴に基づいて検索する点です。

実際の利用手順は以下のようになります。

①キーワード等を用いて公報を検索する。

②検索結果として得られた各公報に対し、それらの図面を一覧提示させる。

③図面の一覧の中から、所望の図面を検索キーに指定して図面検索を実行する。

以上の流れを図1に例示します。

↓　キーワードで検索し図面一覧を提示

↓　検索キーの図面を指定し図面検索実行

図1

図1において、中段の赤い丸で示した図面が検索キーであり、下段にその類似図面が検索されている様子が示されています。

3. 適合性フィードバック

テキスト検索の場合もそうですが、一般に、1回の検索で満足のゆく検索結果が得られることは多くありません。そこで現在の検索結果を、より所望のものに近づける手段が必要になります。

RIPWAYでは、検索結果の図面群の中から、所望のものに近い図面群を検索キーとして追加指定した後、“再検索”を行うことで検索結果をコントロールしてゆきます。図2に、この再検索の様子を示します。

上記の再検索のプロセスは、一般には適合性フィードバックと呼ばれています。すなわち、ユーザ自身による適合性（所望のものか否か）の判断を、システムにフィードバック（教示）することで検索精度を高めてゆく仕組みです。

↓　上記赤丸の図面を指定して再検索

図2

テキスト検索の場合は、最初にキーワードをAND/OR/NOTなどで結合してシステムに与えることで検索要求の内容を90％ぐらいはシステムに伝えることができます。そして残りの10％、つまり、細部を調整してゆくのが一般的な検索プロセスです。

一方、画像検索の場合は、その逆です。すなわち最初に図面を1つ検索キーとして与えたぐらいでは、検索要求の10％ぐらいの内容しかシステムには伝わりません。検索キーとして与える画像を加えることで、次第にユーザの検索意図がシステムに伝わってゆきます。実際には、検索キーの画像群に、なるべく共通に現れる画像特徴こそが、ユーザが所望する内容であるとシステム側が認識できるようになり、その共通特徴の重みを大きくして検索が行われます。

テキスト検索に慣れている一般のユーザにとって、上記の違いを意識することが、画像検索上達のためのコツだと思います。

4. 画像・テキスト・書誌の統合検索

他の多くの画像検索アプリケーションとは異なり、公報図面検索の場合は、公報明細や「図面の簡単な説明」など、図面に付随するテキスト情報も画像検索に利用できます。また、公報に付与されている特許分類コードなどの書誌情報もしかりです。これらの豊かな情報を、画像検索のプロセスでも利用できるようにすることは、適合性フィードバック以外にもコントロール手段をユーザに提供する上で重要になります。

具体的にはRIPWAYでは、キーワード、および、出願人、IPC、出願日など任意の書誌項目を組み合わせた検索条件により、画像検索の結果を絞込むことが可能です。図3にその効果を例示します。

図3の上段は、左上の最初の2つの図面を指定して検索した結果です。一方、図3の下段は、さらにキーワー

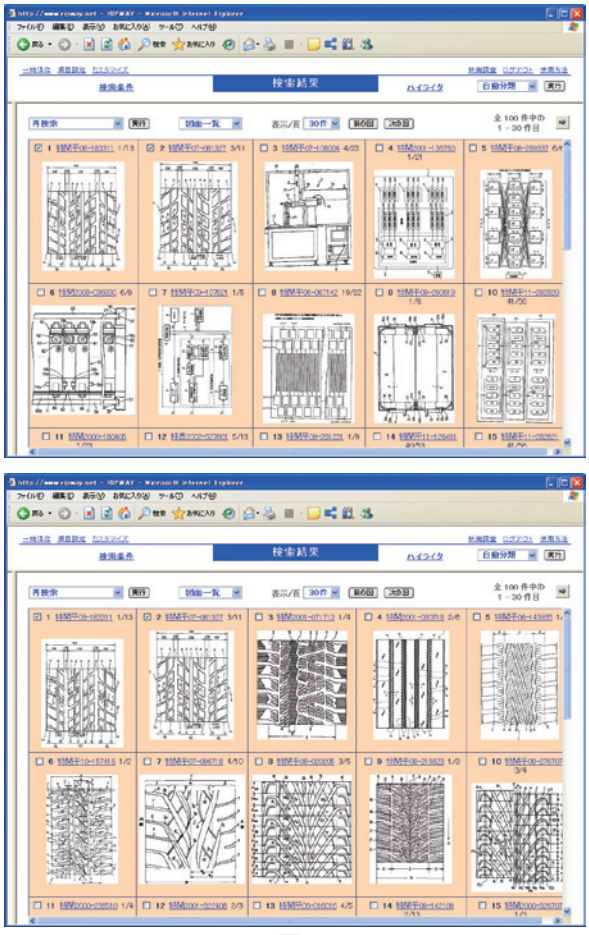

図3

ド条件「タイヤandトレッド」を加えて画像検索した結果です。

検索対象となる公報DBには様々な分野の図面が大量に登録されていますので、検索キーの図面の形状がよほど特殊でない限り、図3上段に見るように、検索意図に沿わない図面が多く検索されてしまいます。そこで、キーワードや書誌項目により、所望する図面の「意味内容」を限定した上で、「形状」の類似性を検索することが必要です。つまり、テキスト・書誌との統合検索は、公報図面の検索において必須機能といえます。

5. 画像の類似性

ここまでで、画像検索の利用イメージを把握していただけたと思いますので、本節では、検索の仕組みについて、ユーザとして知っておいたほうが良い点を述べさせていただきます。

一般に類似画像検索（Content-based Image Retrieval）は、画像の特徴を抽出し、その特徴に基づいて検索および類似度の計算を行います。RIPWAYでは、機械部品、化学式、写真など、図面内容の違いによらず、同じ方法で特徴を抽出します。具体的には、図面を構成する画素（白黒のドット）の分布の偏りやパターンを特徴とします。この特徴は、図面全体に関するものなので、類似性は、図面が全体として似ているか否かで判断されます。つまり、図面間の部分対部分の類似性は考慮されません。

また、上記の特徴は、形状の縦横の伸縮（スケール）には不変ですが、回転に対しては異なるものになります。よって同じ図面でも、例えば左に90度回転されたものとは、類似と見なされません。

以上からRIPWAYで用いている特徴では、部分的に類似、あるいは、回転等のバリエーションには対応できないという問題があります。

にもかかわらず、この特徴を用いる第1の理由は、検索が成功したか失敗したかの判断がユーザにとって容易であるからです。部分的に類似する、回転すると類似する、などの図面を混ぜた検索結果は、多くの場合、ユーザにとって理解が困難なものになります。第2の理由は、検索キーの図面に対し、ユーザが明示的に部分指定したり、回転を施せる機能を設ければ、多くの場合に対応できると考えるためです。

画像検索は、画像特徴というユーザには見えないキーで検索するため、ブラックボックスになりやすい機能です。テキストの概念検索もそうですが、なるべく検索プロセスをホワイトボックス化し、かつ、ユーザによる明示的なコントロールを可能にすべきであるという方針に沿ってRIPWAYの図面検索は設計されています。

6. 検索システム

国内公報（特許・実用新案）については、1993年以降の電子化公報に含まれる全ての図面（選択図以外も含む）および化学式が検索対象です。一方、米国公報については1979年以降に発行された公報の代表図のみが検索対象です。これらの内、公開と登録公報での重複を除くと、実質的には現在、約6000万枚の図面が、検索対象となります。

よって、検索システムとしては、これら大量の図面を実用的な時間で検索できること、公報発行毎の図面データ登録が高速であること、そして、第4節で述べたようにキーワードや書誌との統合検索が可能であることなどが要件となります。

この点、通常の画像検索サービスでは、画像検索機能は画像検索専用のサーバで賄われるため、第4節で述べた統合検索を実現するには、書誌やテキスト用の検索サーバの結果と、画像検索の結果を、アプリケーションレベルで統合することになり、検索スピード上の問題になります。また、システム全体が複雑化し、導入や運用のコスト上の問題も生じがちです。

RIPWAYでは弊社製品である検索サーバTRMeister（tm）（ティーアールマイスター）が検索エンジンとして用いられています。この検索サーバでは単体で、RDB機能に加え、テキストおよび画像の両方の検索を実現することができます。検索サーバ自体は汎用目的で開発され、文書管理や図書館システム等でも利用されていますが、登録と検索のスピード、多言語対応（日英、欧州5ヶ国語、中国語）、概念・類似検索精度などの面から、知財分野の検索エンジンにも適しています。サーバレベルでテキスト、書誌、画像の検索が統合されるので、高速な画像検索を実現することが可能です。

7. 図形商標の類似検索

知財分野における画像検索という観点からは、特実の公報図面に加え、図形商標、意匠も重要な要素ですので、最近、弊社で試作を終えた図形商標の類似検索システムについて、簡単にご紹介させていただきます。

図形商標の類似検索の利用法や仕組みとしては、前述の公報図面のものとほぼ同様ですが、以下の機能も実現されています。

①色を捨象した検索
②画像の加工・変換
③部分検索

公報図面とは異なり、図形商標の画像はカラー画像です。しかし図形商標の類似性判断では、色の違いに依らない、形状のみによる検索が必要となります。図4には、左上赤枠内の図形商標（検索キー）に対して、色の違いによらず、類似図面が検索される様子が示されています。

図4

また、指定した図形商標について、部分の切り取りや、回転、左右反転などの画像変換を行い、その結果を検索キーとすることも可能です。図5は、図形商標上で、赤枠の矩形領域をマウスで指定し、切り取っている様子を示しています。

図5

さらに、図6に示す検索例のように、左上赤枠の検索キーの図形を、部分として含む図形を検索することが可能です。前述のように、こうした部分検索の検索結果はユーザにとって理解しにくい場合が多いので、本システムには、システムが類似していると判断した箇所を、検索結果の各図形上に赤の矩形で表示する機

能が組み込まれています。

図6

このように、検索キーの図形に対する加工や変換、検索対象の部分検索、適合性フィードバック、ウィーン分類コードとの統合検索などを組み合わせ、ユーザが様々な観点から検索結果をコントロールできることが大事であると思います。このことは、図7に示すような、複雑な図形商標の検索を行う場合に、より重要となります。

図7

8. おわりに

以上、特許情報サービスにおける画像検索についてご紹介・ご説明して参りましたが、最後に今後の可能性について述べさせていただきます。

公報図面の検索については、現在は前述のように、図面内容の違いによらず、一律の方法で特徴を抽出しています。しかし公報にはIPCやFターム、図の説明などの豊かなメタ情報が付与されていますので、こうした情報を基に、図面の内容区分を同定し、その別に応じた特徴を用いるようにすることで、検索精度を更に向上させることができると思います。このことは図形商標におけるメタ情報の利用についても当てはまります。

また本稿では意匠の検索については触れませんでしたが、基本的には図形商標と同様の枠組みで実現できると思います。この場合、意匠特有の立体情報に関して、6面図毎の画像検索結果を統合すること、また、部分意匠に関しては、本稿で述べた部分検索の機能を適用することが課題であると考えております。

謝辞：

山田和彦様をはじめ（財）日本特許情報機構の特許情報研究所の方々に、図形商標・意匠の審査等に関しご教示いただいたことを感謝いたします。また特許庁技術懇話会委員の方々には、この機会を賜りましたことを感謝いたします。

Profile

伊東 秀夫（いとう ひでお）

1985年　株式会社リコー入社
1986年～1999年
　　機械翻訳など自然言語処理の研究開発に従事
2000年　東京工業大学大学院　博士後期課程修了（工学博士）
2000年～2004年
　　TREC,NTCIR参加など情報検索の研究開発に従事
2004年～特許情報サービスRIPWAYの事業化支援など、現在に至る