

知財戦略を支える「XML特許統計サービス」

企業の知財戦略が重要度を増すにつれ、競合他社の特許出願に関するデータの統計処理に対するニーズが高まってきている。しかしながら特許情報の迅速な統計処理は、データ件数が膨大であるため容易ではない。本稿では、特許情報サービスの動向と、インターネットで利用できるNRIサイバーパテントのXML特許統計サービスについて紹介する。

特許情報サービスを取り巻く環境

近年、特許情報サービスを取り巻く環境が急速に変化している。

そのひとつは、企業が知財戦略を重視するようになってきていることにともなって、特許情報サービスに対する企業のニーズが変化していることである。従来、特許情報サービスに求められていたのは、自社が特許を取得しようとする際に必要となる、先行技術や無効資料の調査をインターネット上でサポートする、いわゆる調査支援ツールとしての機能であった。しかし、いまではそれに加えて、自社の知財戦略の意思決定を支援するために、競合他社の特許出願の動向などを多面的に調査・分析する統計機能へのニーズが高まっている。

もうひとつの大きな変化は、特許データの形式の変更である。特許庁は1993年から特許情報をSGML（標準マークアップ言語）形式の電子データで提供してきたが、これが2004年からXML（拡張可能なマークアップ言語）形式へと変更された。これにより、XMLをベースとした特許情報サービスのアプリケーション開発が急速に進むと考えられる。

取得が困難な特許情報統計データ

競合他社がこれまでにどのような技術分野で特許出願を行っているのかなどの統計処理ができるようになると、たとえばどの企業と技術提携したらよいかといった判断も容易となる。また、どの研究者が発明によって企業に貢献しているかも把握できるので、ヘッドハンティングの際の資料としてデータを活用することも可能になる。このように、特許情報の統計処理により、特許情報サービスの有用性はますます高くなるわけである。

しかしこのような特許情報の統計処理は、データの累積件数が数百万～数千万件のオーダーとなるため、そのサービスをインターネット上でリアルタイムに提供することはきわめて困難である。そのためこれまでは、特許情報に関する統計データを作成しようとすれば、調査支援機能によって得たデータを表計算ソフトなどで加工するといった手作業が必要となっていた。

特許データの形式変更への対応

特許データの形式変更も、特許情報サービスのあり方にとって大きな意味をもってい

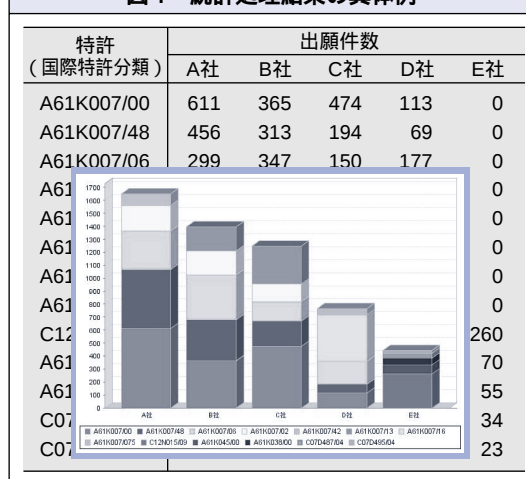
る。近年はXMLをベースにしたアプリケーション開発がさかんに行われており、また一般に大量データの統計処理にはXMLが向いている。そのため、今後は特許情報サービスでもXMLが業界標準となることは間違いない。しかし、特許データの形式変更は、特許データが電子化されて以来初めてのことであるため、特許データを取り扱うシステムへの影響はきわめて大きいものがある。特許制度の性質上、過去のSGML形式の特許データも特許情報サービスの対象としなければならないため、SGML世代のデータとXML世代のデータをシステム上で横断的に取り扱えるよう対応しなければならなくなったのである。

NRIサイバーパテントの「XML特許統計サービス」

以上のような背景を踏まえて、NRIサイバーパテントでは、1993年以降の約400万件のSGML形式の特許データのすべてを特許庁の仕様に応じたXML形式へと変換した。

これにより、SGMLで提供されていた特許データはすべてXMLとして横断的に取り扱えることになり、XMLをベースとした統計処理アプリケーション開発の土台を築くことになった。そしてXML特許データはすべて高速アクセスが可能なサーバーの内部メモリーに展開され、即時に検索・集計ができるXML特許統計サービスを提供することが可能となったのである（図1参照）。

図1 統計処理結果の具体例



本サービスでは、統計処理の対象とする書誌的な事項を1つまたは2つ指定すると、関連する特許出願件数を集計できる。また目的に応じてアウトプットの形式を選択することも可能である。さらに、母集団として企業名および発明の属する技術分野を指定した上で、1次元または2次元でのクロス集計を行うことも可能である。集計対象は、個別に指定する方法のほか、出願件数の上位を自動抽出（ランキング）することも可能になっている。なお、XML特許統計サービスはインターネット上で利用できるようになっており、自社の得意とする技術分野で競合他社がどのような取り組みをしているかなどのデータを即時に取得できる。

知財戦略のあり方が企業の将来を左右するような情勢のなかで、XML特許統計サービスは知財戦略に係る決定をスピーディーに行うための強力なツールとなるであろう。 ■