

ネットワーク上の 文書テキスト検索・利用技術

鈴木克志*
高山泰博*

要 旨

文書の電子化が進み大量の情報がネットワーク上に存在する現状は、情報過多又は情報過負荷(Information Overload)と呼ばれる。このような状況においては、計算機に結合されたネットワーク全体の文書資源を活用する文書の検索機能が、IT(情報技術)における重要な課題となってきた。

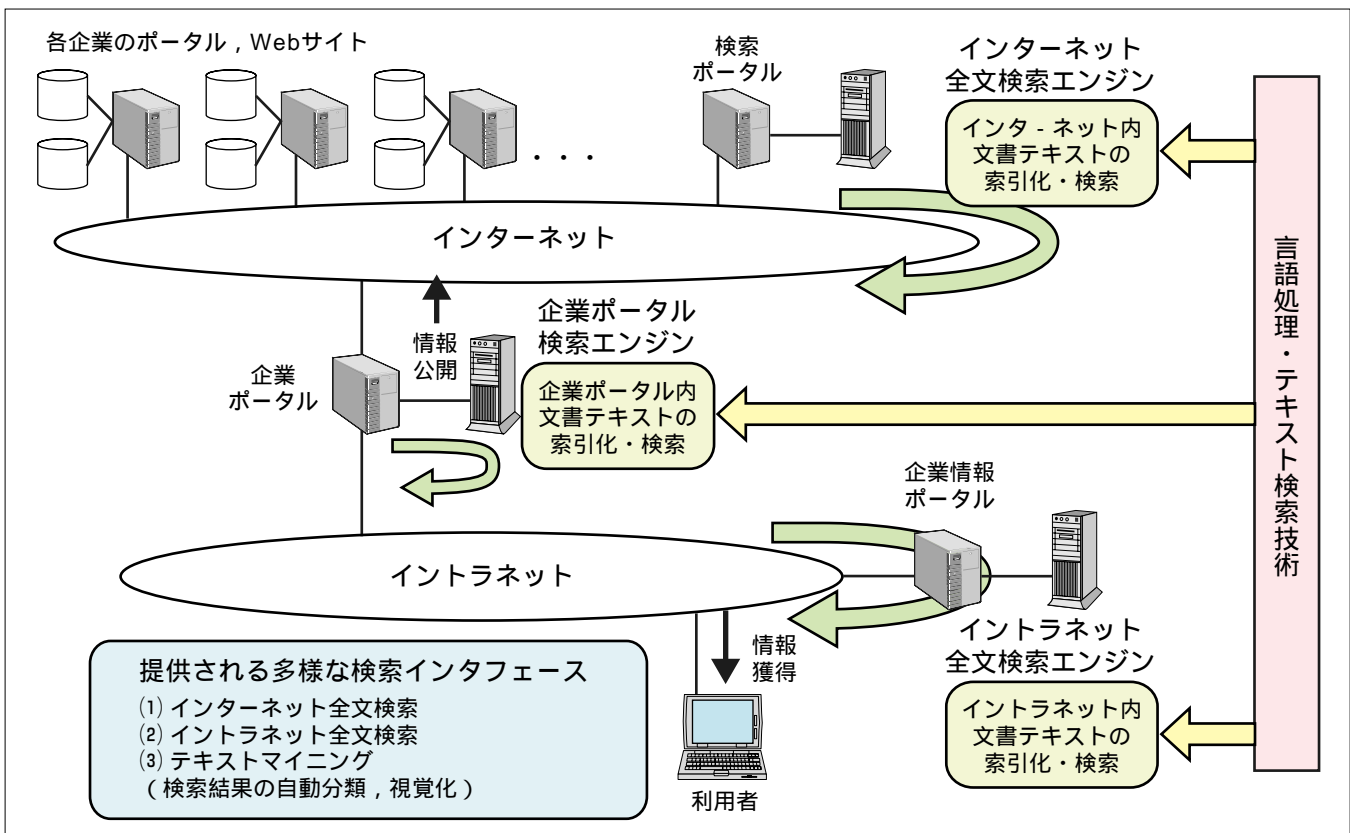
一方、ヒューマンインタフェースの観点から見ても、文書の検索機能は、ネットワークと人間を介するインタフェースの重要部分を占めていると考えられる。

文書検索ソフトウェアは、利用者から見て計算機システム全体のバックエンドに位置するプロセッサである。しかし、近年、バックエンドに位置付けられる検索エンジンの呼出しが、フロントエンドでの操作機能として重要な位置を占めるようになった。すなわち、最近では、オペレーティングシステムでの不可欠機能としてフロントエンド

タフェースに全文検索の入力窓が搭載されるようになり、Webの画面上に検索入力窓を設けているポータルも多くなっている。

さらに、検索結果を業務で利用する新しい応用が生まれてきており、そこでは、業務に応じた検索結果の後処理方式やフロントエンドインタフェースにおける検索結果の表示方式などの工夫が行われている。

そこで、本稿では、ネットワーク全体の文書資源を活用するための文書テキスト検索技術の重要性に着目し、文書テキスト検索の基本原則、検索方式の分類、及び重要な基盤技術としての形態素解析について述べる。さらに、検索ポータル、企業ポータル検索エンジン、インターネット情報収集、及びテキストマイニングなど、三菱電機の言語処理技術やテキスト検索技術によって実現可能となった先進の応用事例について紹介する。



ネットワーク上の文書テキスト検索ソリューションの概念

先進の言語処理技術とテキスト検索技術により、従来困難であったネットワーク上の大量文書の活用が正に可能な時代を迎えた。利用者は、種々の検索エンジンが提供する検索ソリューションによって、多様な検索機能を利用することができる。検索ポータルは、インターネット全文検索エンジンによって大量のWeb文書の検索機能を提供する。また、企業情報ポータルは従業員に企業内のテキスト検索を提供し、企業ポータルはポータル訪問者にコンテンツへの検索機能を提供する。これらの検索結果は、目的に応じたテキストマイニング機能によって業務に活用される。