

インターネットを利用した全文検索システム

3P-8

○鈴木健 今村泰介 山口文雄 杉本未紀

(株)東芝 ソフトウェアプロダクト部

1. はじめに

近年、インターネットの普及が著しく、様々な情報が溢れている。このような情報洪水の中での確かつ迅速に目的の情報を引き出す、あるいは、提供するための機能が、情報を活用するうえで不可欠となっている。そのため、情報検索の機能の早急な実現が望まれていた。

一方、当社では、全文検索システムとして、高速全文検索システム FineSearch™（以下、FineSearch™と呼ぶ）が既に製品として開発されている。FineSearch™はAPI (Application Programming Interface) を有する検索エンジンと、対話的に検索を行なうGUIを持つクライアントで構成される。

そこで、このAPIとWWW (World Wide Web) サーバのCGI (Common Gateway Interface) 機能を組み合わせることにより、簡単に、WWWによる情報検索サーバ構築ソフトウェア FineSearch/Web™を実現できた。本稿ではその概要を報告する。

2. 機能概要

FineSearch/Web™により、既に構築されているWWWサーバ内にあるHTML文書の検索サービスや社内における報告書の検索サービスなどを行なうことができる。

FineSearch/Web™には、次の機能がある。

(1) 検索システムを構築するユーザ（以下、システム管理者と呼ぶ）用機能

・検索画面作成機能

実際に検索を行なうための検索画面となるHTML文書を作成する機能。検索システム構築時に、この機能を用いて検索画面を作成し、検索システムとして提供する。

(2) 実際に検索システムを利用するユーザ（以下、システム利用者と呼ぶ）用機能

・検索機能

ブラウザの検索画面に入力されたキーワードなどをもとに検索し、結果の一覧をブラウザに表示する機能。検索はフリーキーワードで行なえ、類語検索、分類（カテゴリ）別の検索も行なえる。

・該当文書の内容表示機能

検索結果の一覧の中から指定された文書の内容を、ブラウザに表示する機能。文献中の検索文字列にタグをつけておき、順次表示することもできる。

3. 動作の仕組み

システム利用者用の機能である検索、該当文書の内容表示という一連の流れにおける検索サーバへ

Full text search system for Internet

Ken Suzuki, Taisuke Imamura, Fumio Yamaguchi, Miki Sugimoto

Package software products dept., Toshiba corporation

の従来のアクセス方法を図1に示す。従来、このような一連の流れでは、ひとつのクライアント用モジュールを起動し、始めに検索サーバへ接続し（図1中の2）検索サーバに接続した状態でユーザとやり取りを行ない（図1中の3～10）最後に切断する（図1中の12）という流れになる。

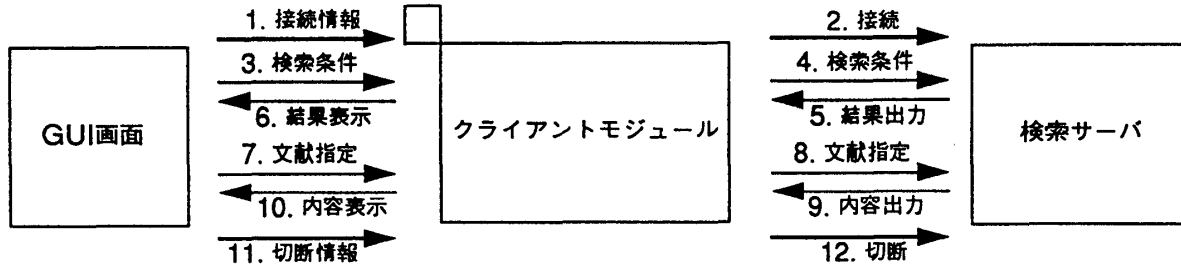


図1 従来 (FineSearch™) の処理の流れ概略図

一方、WWWサーバと連携させるにはWWWのCGIを利用する。CGIはその性格上、WWWサーバからモジュールへの情報受け渡しはモジュール起動時にしか行なえず、モジュール起動中にインタラクティブな受け渡しを行なうことができない。したがって、図2に示すように、各機能ごとに異なるモジュールを起動させ、そのたびに検索サーバに接続/切断（図2中の2, 5および8, 11）している。このとき、一連の流れにおける必要な情報は、図2中の6で出力するHTML文書内に記述しておき、内容表示モジュールをCGIで起動するとき引数として渡すことで実現している。

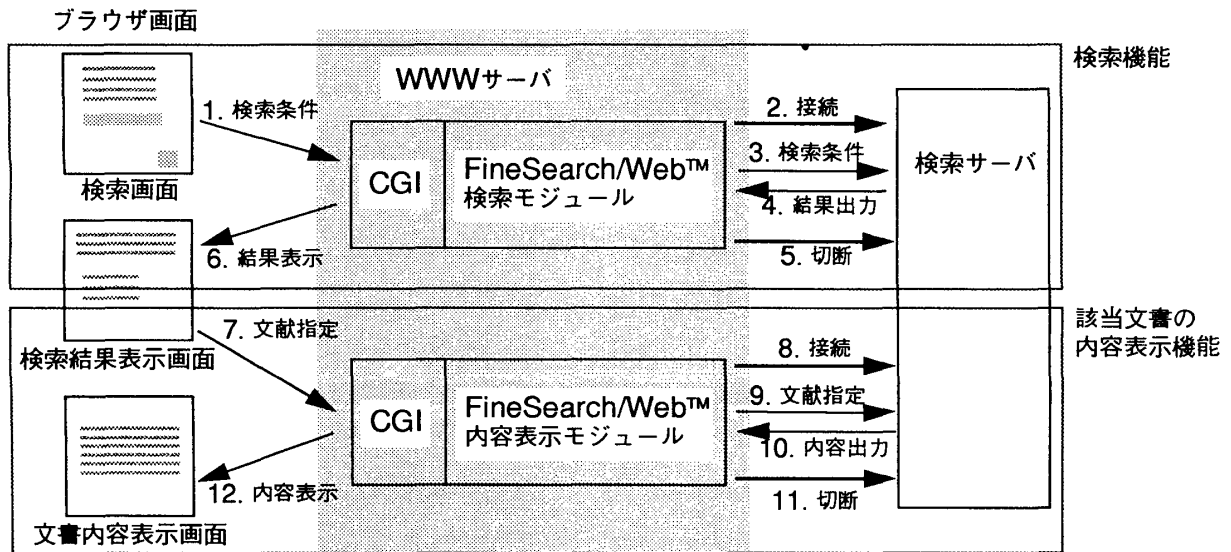


図2 FineSearch/Web™の処理の流れ概略図

4. おわりに

WWWによる情報検索サーバ構築ソフトウェアの開発にあたり、WWWサーバのCGI機能を利用し、既存のFineSearch™のAPIを使用することにより、実現することができた。そして、この実現により、FineSearch™による全文検索機能を用いて管理された文書について、WWWブラウザからキーワード検索し、該当する文書を表示することができるようになった。このように、ネットワークをまたがったグローバルな全文検索システムの構築が可能となった。