

# 大規模文書情報システム用 テキストサーチマシンの研究

加藤 寛次   藤澤 浩道   大山 光男   川口 久光   畠山 敦

(株)日立製作所 中央研究所

インデックス情報を用いないスキャン型のフルテキストサーチを実現するための技術課題を明らかにすると共に、これを解決する方法として、①同義語・異表記検索方式、②並列駆動式集合型磁気ディスク装置、③高速多重文字列照合方式、④サロゲート型検索加速方式、⑤パイプライン型複合条件判別方式、⑥ビルディングブロック式システムアーキテクチャを採用したテキストサーチマシンを提案する。そして、これらをインプリメントしたプロトタイプTSM-Iについて報告する。本プロトタイプは、12台の小型磁気ディスク装置と1台のサーチエンジンから構成され、等価的に100MB/sのシステム検索速度を得ることができた。

## A study of a full-text search machine for large Japanese text database

Kanji KATO   Hiromichi FUJISAWA   Mitsuo OHYAMA  
Hisamitsu KAWAGUCHI   Atsushi HATAKEYAMA

Central Research Laboratory, Hitachi, Ltd.  
KOKUBUNJI, TOKYO 185, JAPAN

A full-text search machine requiring no indexing operation is presented that is for large Japanese text database. The machine consists of N search units operating in parallel. Each search unit has a high speed search engine to scan the source text stream searching for many query terms including synonyms and their spelling variants. To get a fast text stream, each unit has K disk drives to read text data in parallel. To accelerate the speed further, it adopts a two-stage surrogate method. A single-unit prototype machine has developed which has 12 5-1/4" disk drives. The total equivalent scan rate of 100 MB/s per unit has been attained.

## 1. はじめに

近年、ワードプロセッサやパーソナルコンピュータ、ワークステーションなどの普及拡大に伴い、作成される文書情報も急速に増加してきており、近い将来膨大な量に達するものと予想されている。このため、大量の文書情報を一般のユーザが簡単に蓄積・検索できる文書情報検索装置に対する要求が高まりつつある。また、既存システムにおいても、文書データベースの大規模化に伴う絞り込み性能の低下や、技術文献データベースでの技術用語の目まぐるしい変遷に起因する検索精度の低下などが大きな問題となってきた。

こうした要求や問題に応えるものとして、インデックス情報を用いない自由な言葉による検索を目的としたフルテキストサーチ(自由語全文検索)技術の研究を行なってきている。

このインデックス不要型のフルテキストサーチを実現するためには、テキストデータを蓄積ファイルから高速に読み出し、この中から所望の言葉をオンザフライで探し出す技術のみならず、サーチ速度を等価的に高め検索時間を実用領域に持ち込む検索加速技術が必要となる。

本研究では、テキストデータを高速にスキャンしサーチするために、テキストデータを高速にアクセスできる並列駆動式の集合型磁気ディスク装置を開発すると共に、指定された言葉だけでなく、その同義語や異表記語をも含めて一括して探索する文字列照合専用プロセッサ(サーチエンジンSEと呼ぶ)を開発した。更に、検索時間を実用範囲に収めるために、フルテキストサーチ速度を等価的に加速するサロゲート型の多段検索絞り込み方式を開発した。そして、これらのアーキテクチャを一つのシステムとしてまとめたプロトタイプ(テキストサーチマシンTSM-Iと呼ぶ)を試作した。

ここでは、はじめに文書検索における問題を概観し、その課題に対するアプローチとして提案しているフルテキストサーチ方式について説明する。そのあと、これをインプリメントしたテキストサーチマシンTSM-Iのアーキテクチャとその処理内容について述べる。最後に、その試行実験結果について報告する。

## 2. 文書検索における課題

### 2.1 従来の文書検索における問題

#### (1) 統制語による索引付けの問題

従来、文書データベースの検索システムにおいては、キーワードや分類コードによる検索方式が用いられてきた。[1]

データベースに文書情報を登録する際には、インデクサーと呼ばれる専門家がその登録文書の本文を

読み、その内容を表す語彙を統制語辞書(シソーラスと呼ぶ)から選び出して、これをキーワードとしている。これを索引付け(インデキシング)と言い、このキーワードを基にインバーテッドファイルと呼ぶインデックスが作成される。そして、このインバーテッドファイルを用いて検索が行なわれる。もし、このインデキシングが適切に行われないと、データベースへの登録後に検索できないことになる。

検索の際にも同様に、シソーラスから検索内容を表す語彙を拾い出して、これをキーワードとして検索を行う。もし、シソーラスにないキーワードを指定すると、目的とした文書が呼び出せないことが生じる。あるいは、不適切なキーワードを指定すると、不要な文書(ノイズと呼ぶ)まで検索してしまうことになる。

したがって、インデキシングのみならず検索においても、文書の技術分野に精通し、シソーラスにも熟知した検索の専門家が必要となる。

また、この文書登録時のインデキシングは非常に煩雑な作業を伴う上に、インバーテッドファイルの作り直しも必要となってくる。更に、この作成には多大の処理時間を要するため、システム運用上の大きな課題ともなっている。したがって、新たに文書を追加登録する際には、バッチ処理によってかなりの量をまとめて登録処理することになるため、最新の情報が検索できないという問題も生じてくる。

#### (2) データベースの保守の問題

また、技術そのものが年月と共に進歩、変化するために、シソーラスにおける分類体系やキーワードも時代と共に陳腐化してしまうという問題がある。このため、キーワードや分類体系を常に更新しなければならないということになるが、登録済みの膨大な文書を読み直して再インデキシングし、インバーテッドファイルを作成し直すのは、大規模なデータベースの場合、事実上困難となる。

#### (3) データベースの大規模化の問題

また、データベースが大規模化するに従い、シソーラスに記述された統制語だけでは文書の内容を十分詳細に記述できないため、キーワードで検索しても絞り込めなくなってきた。この中から目的とする文書を見つけ出すためには、それらの内容を直接読むしか方法がなく、これが検索効率上の大きな問題となっている。

このシソーラスを用いたインデキシングに基づく現状の検索方式の問題に対して、自動抄録や自動インデキシングの試みがなされてきているが、日本語の場合その言語的な困難性から、やはり種々の辞書を必要とするため上記の本質的な問題の解決に至っていない。[2],[3],[4]

## 2.2 フルテキストサーチの課題

こうした問題に対する根本的解決方法として、検索者が自由なキーワード(自由語とも呼ぶ)に基づいて文書の本文を直接参照して内容を検索できる全文検索(フルテキストサーチとも呼ぶ)システムが提案されている。[5]

文書の本文を直接参照して内容を検索するこのフルテキストサーチでは、シソーラスに基づくインデキシング及び検索が不要となるため、自由な言葉で検索が行えるようになり、呼び出し率が高くノイズの低い検索が可能となる。また、キーワードを任意に設定できるため、データベースが大規模化しても絞り込み性能が低下するという問題を回避できることになる。更に、インバーテッドファイルを用いないため、簡単に文書登録が行え、かつ登録した文書は即座に検索できることになる。また、年月の経過に伴うシソーラスの更新及びこれに付随する再インデキシングも不要となるため、常にその時点の必要な観点から検索ができることになる。

しかし、以上述べたフルテキストサーチを実用的な規模のテキストデータベースに適用しようとする時、下記のような問題が発生してくる。

### (1) 検索時間の問題

先ず第一に、検索時間の問題である。

#### (a) 文字列照合速度の問題

これには先ず、指定キーワードを探索するための文字列照合処理を高速化する必要がある。そのため的高速化アルゴリズムや、専用ハードウェアが提案され、様々な方式の研究が行なわれてきている。[6],[7],[8],[9],[10]

#### (b) 照合文字列数の問題

しかし、自由語を用いたフルテキストサーチでは、しばしば検索者が指定したキーワードと、テキスト本文中に記述されている言葉の間に、同じ意味を表していても表現に食い違いが生ずることがある。このような場合には、指定したキーワードと異なる表現形態を持つ文書が検索漏れとなることになる。こうした言葉の例として、同義語や異形語(異表記語あるいは単に異表記とも呼ぶ)などがある。

同義語の例としては、「計算機」に対して「電子計算機」や「電算機」、「Computer」などが挙げられる。また、異形語の例としては、「コンピュータ」に対して「コンピユータ」や「コンピユータ」、「コンピユータ」、「コンピユータ」、「コンピユータ」などが、「Computer」に対しては「computer」、「COMPUTER」などが挙げられる。

この問題に対処するためには、検索装置の内部で

ユーザが指定したキーワードの同義語及び異形語を自動的に生成して検索を行う必要がある。これには、数百にも及ぶ異形語や同義語を照合速度を落すことなく、一括して照合し得る文字列照合回路が必要となる。

#### (c) テキストファイルの読出し速度の問題

一方、文字列の照合処理をどれだけ高速化しても、テキストデータを格納する蓄積ファイルの読出し速度が低くては高速なサーチは不可能である。したがって、テキストデータを納める蓄積ファイルの読み出し速度を文字列照合回路の照合速度と同程度にまで高めることも必要となってくる。

#### (d) データベースの大容量化の問題

しかし、蓄積ファイルの読み出し速度と文字列照合回路の処理速度を高めたとしても、大容量のテキストデータベースをスキャンする場合、検索時間がユーザの耐えられないほどになってしまう可能性がある。したがって、データベースが大容量化した場合でも、フルテキストサーチの質を損なうことなく、検索時間を実用上許容し得る数秒台に納めるための加速手段が必要となってくる。

#### (e) 複合条件処理速度の問題

また、フルテキストサーチ特有の機能としてキーワード間の文字距離を指定した近傍条件検索や、キーワードの文脈上の位置関係を指定した文脈条件検索がある。ユーザから与えられる検索式には、このほかに一般的な論理条件(論理和や論理積など)なども含まれ、これらが複合的に結合された形で検索式が与えられる。

このような近傍、文脈、論理条件などの処理は非常に煩雑であるため、処理に時間が掛かるという問題がある。したがって、高速なフルテキストサーチを実現するためには、複合条件の判別処理を高速に行うことが必要となる。

### (2) システムアーキテクチャの問題

#### (a) データベースの拡張性問題

さらに、テキストデータベースに文書データが逐次登録されて行くと、ある時点で蓄積ファイルの容量が満杯に達してしまう場合が生じる。こうした時にも、それまで蓄積したデータを損なうことなく、システムの蓄積容量を拡大できることが必要となってくる。

#### (b) データベースの大規模化の問題

また、テキストデータベースの容量が大規模化してきた場合にも、検索時間を低下させることなく、これに対応できるようなシステムアーキテクチャが必要となってくる。

## 3. フルテキストサーチ方式

上記の課題に対して、本テキストサーチマシン

TSMでは以下のようなフルテキスト  
サーチ方式を採った。

### 3.1 記述・表記違いの検索漏れ防止

ここでは、図1に示すように、ユーザが指定したキーワードの同義語と異表記語をシステム内部で自動的に生成して、これらをまとめてキーワードとして検索する同義語・異表記検索方式を採用した。同義語展開は同義語辞書を参照する方式とし、異表記展開にはルール変換方式を採用した。これにより、ユーザが指定したキーワードとテキスト中に記述された言葉の間の食違いにより生ずる検索漏れを防止、呼び出し率を高めることが可能となった。

### 3.2 高速なスキャン型検索処理の実現

フルテキストサーチの最も重要な課題となる高速化については、以下のアプローチを採った。

#### (1) テキストデータの高速な読出し

まず、テキストデータを納める蓄積ファイルの読み出し速度を文字列照合回路の処理速度と同程度まで高めるために、図2に示すような複数台の磁気ディスク装置を並列に並べた集合型の磁気ディスク装置（マルチディスクユニットMDUと呼ぶ）を開発した。このマルチディスクユニットMDUを構成する各小型磁気ディスク装置は、それぞれ独立にアクセス制御ができる構造とし、複数文書ファイルのシークンシャルな読出しのみならず、ランダムな読出しも高速に行なえるようにした。

#### (2) 高速な多重文字列サーチ

多数の同義語や異表記語を照合速度を低下させることなく探索するため、テキストのただ1回の走査で約一千語のキーワードを一括してサーチできる有限オートマトン型の多重文字列照合方式を開発し、専用プロセッサ（サーチエンジンSEと呼ぶ）としてハードウェア化した。

#### (3) サーチ速度の加速

並列に並べた複数台の小型磁気ディスク装置で構成されるマルチディスクユニットMDUとサーチエンジンSEによりテキストデータを高速にスキャンできるようにすると共に、更にスキャン速度を上げ検索時間を実用上許容し得る数秒台にまで短縮するために、図3に示すよう

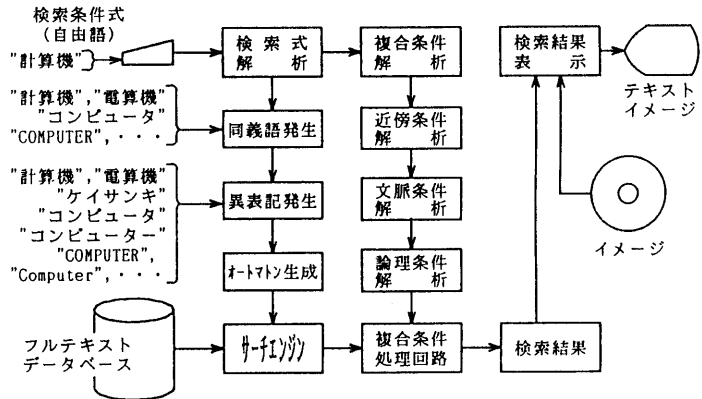


図1. 自由語検索の処理方式

なサロゲート型の検索加速方式を開発した。

#### (a) 文字成分表サーチ

この検索加速方式は、2段階のプリサーチで構成されている。その第1段階目のプリサーチは、予めテキスト中に含まれる文字を抽出し、ビット情報で表しておいたテーブル（文字成分表と呼ぶ）をキーワードを構成する文字毎にサーチし、全ての文字を含む文書を割り出す処理からなる。これを文字成分表サーチ（ビットサーチ）と呼ぶ。これは、キーワードを含み得ない文書を切り落とす処理である。これによりキーワードを含まない無駄な文書をスキャンせずに済むことになる。

#### (b) 凝縮本文サーチ

第2段階目のプリサーチは、予めテキスト本文の中から助詞や接続詞などの付属語を削除すると共に、繰り返し表れる単語を排除したテキストファイル

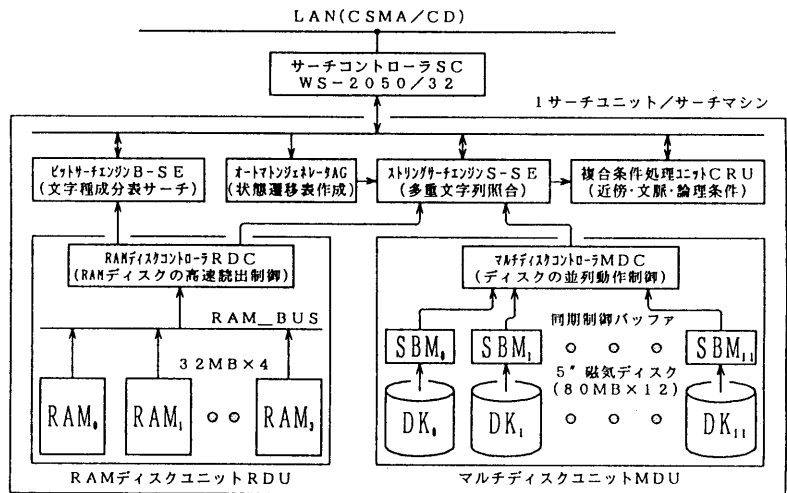


図2. 試作テキストサーチマシンTSM-Iの構成

(凝縮本文と呼ぶ)をスキャンし、指定したキーワードが記述されている文献だけを抽出する処理からなる。これを凝縮本文サーチと呼ぶ。ここでは、キーワードを単語として含む文書だけを拾い出すことになる。

(c) 本文サーチ

この2つのプリサーチでは、先ず最初に与えられたキーワードを1文字単位に分解して文字成分表サーチを行なう。次に、これでヒットした文書に対して、凝縮本文サーチを行なう。最後に、凝縮本文サーチで絞り込まれた文書に対して、本文そのものをスキャンし複合条件を含めて検索条件に合致するものを抽出する。したがって、この2段階のプリサーチを通して絞り込まれた文書についてのみテキスト本文を磁気ディスク装置から読み出してスキャンすることになるため、等価的に非常に高速なフルテキストサーチが実現できることになる。

(4) 複合条件の高速判別処理

フルテキストサーチ特有の木目細かな検索を実現するための近傍、文脈、及び論理条件からなる複合条件検索を高速に実行できるようにするため、サーチエンジンSEからの照合結果を受けて近傍条件と文脈条件並びに論理条件をこの順に一貫して処理するパイプライン型の複合条件判別方式を開発した。この方式では、図4に示すように、サーチエンジンSEの照合結果出力として照合キーワードの識別情報のほかに、照合位置情報を付加して出力するようにし、この位置情報を基に先ず近傍条件、次に文脈条件を、最後に論理条件を判別する処理構成となっている。これら3つの処理は完全にパイプライン化されたアルゴリズムに基づくため、マルチプロセッサ化により容易に高速化が可能となる。

3.3 システムの拡張性の確保

テキストデータベースの容量が大規模化した場合にも、検索時間を低下させることなく対応できるように、図5に示すようなビルディングブロック式のアーキテクチャを採用した。

すなわち、複数の小型磁気ディスク装置で構成されるマルチディスクユニットMDUとサーチエンジンSE、複合条件判別ユニットCRUをまとめて1つのサーチユニットSUとし、これを複数個並列に並べ、この上位にこれらを制御するサーチコントローラSCを設けて1台のテキスト検索マシンTSM-1とするビルディングブロック式のアーキテクチャを採った。このような構成

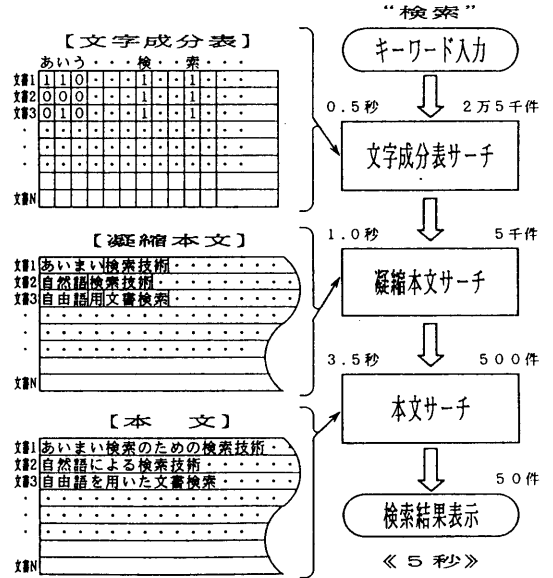


図3. サロゲート型検索加速方式

にすることによって、データベースの増加に際しては、マシン内のサーチユニットSUの増設によって性能を落すことなく対処できることになる。また、複数台のテキスト検索マシンTSMをLANを介して併設し、これらを並列に駆動することによって、更に大規模なデータベースをも容易に構築できるようになると共に、複数のユーザがこの大規模なテキストデータベースを共有し、同時に検索を実行することも可能となる。

4. プロトタイプの試作

以上説明したフルテキストサーチ方式に基づくテキスト検索マシンのプロトタイプTSM-1の試作を行なった。(図6参照)

今回試作したプロトタイプは、12台の小型磁気

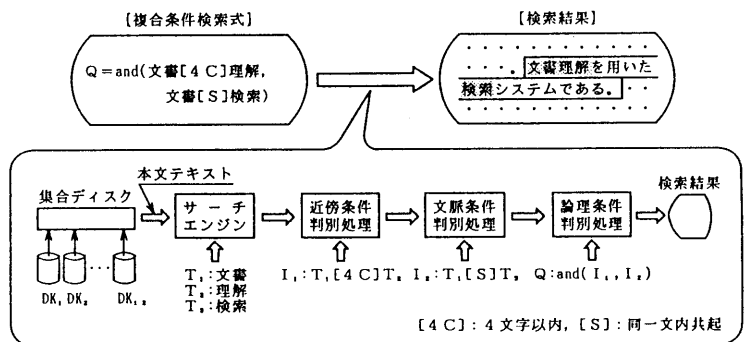


図4. パイプライン型複合条件判別方式

ディスク装置と1台のサーチエンジンSEから成る1サーチユニットSUであり、これはサーチコントローラSC(ワークステーション2050/32)を介してLAN(CSMA/CD)へ接続される構成となっている。

#### 4.1 構成及び仕様

以下、このテキストサーチマシンTSM-Iの構成と仕様について、図2及び表1を用いて説明する。

12台の小型磁気ディスク装置から構成されるマルチディスクユニットMDUには、文書の本文のほか、凝縮本文と文字成分表が格納される。各磁気ディスク装置からのテキストデータの読出しはマルチディスクコントローラMDCで制御される。マルチディスクユニットMDUの容量は960MBであり、平均20KBの文書が2万5千件格納できる。

RAMディスクユニットRDUには凝縮本文と文字成分表がロードされ、RAMディスクコントローラRDCの制御により高速な読み出しが行なわれる。RAMディスクユニットRDUの容量は128MBである。

ビットサーチエンジンB-SEでは文字成分表のサーチが、ストリングサーチエンジンS-SEでは凝縮本文及び本文のサーチが行なわれる。複合条件処理ユニットCRUではストリングサーチエンジンS-SEの照合出力に基づき近傍条件、文脈条件、及び論理条件の判別処理が行なわれる。

#### 4.2 動作の概要

以下、このハードウェアの動作を概説する。

まずユーザが与えた検索条件式は、LANを介してサーチコントローラSCに受け取られ、ここで検索タームの抽出と複合条件式の解析が行なわれる。更に、検索タームに関しては同義語及び異表記語の展開処理がなされ、これらが新たな検索タームとしてビットサーチエンジンB-SEとオートマトンジェネレータAGへ転送される。これを受けて、オートマトンジェネレータAGは受け取った検索タームを照合するための状態遷移テーブルを作成し、これをストリングサーチエンジンS-SEへ設定する。一方、複合条件に関しては、近傍条件、文脈条件、及び論理条件に分解され複合条件処理ユニットCRUへ送られる。

このようにして、各部への制御情報の設定が終了したら、まずビットサーチエンジンB-SEとRAMディスクコントローラRDCに起動を掛け、RAMディスク上の文字成分表をサーチ第1段目の文字成分サーチを行なう。次に、ここで絞り込まれた文書について第2段目の凝縮本文サーチを行なう。凝縮本文サーチはストリングサーチエンジンS-SEとRAMディスクコントローラRDCに起動

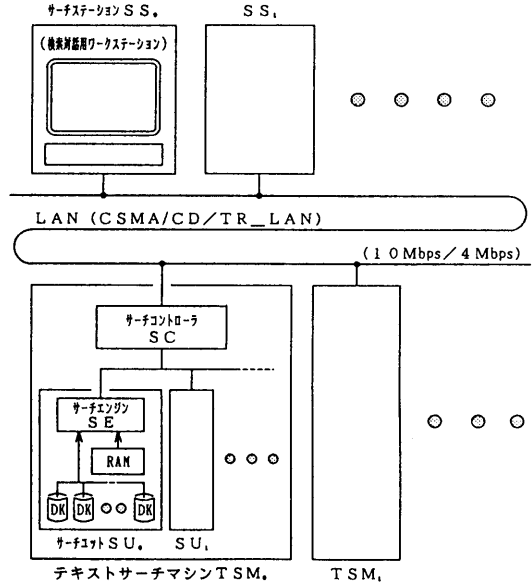


図5. 文書情報システムの構成

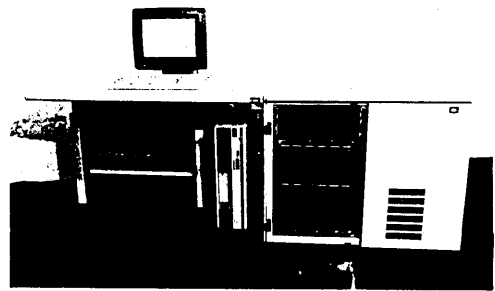


図6. プロトタイプの外観

表1. テキストサーチマシンTSM-Iの仕様

分類	項目	目標仕様
蓄積容量	磁気ディスク容量	960MB
	RAMディスク容量	128MB
	文書件数	25K#
	文書容量/件	20KB
	データベース数	1種
検索速度	検索時間	5秒
	サーチエンジン速度	20MB/s
	磁気ディスク速度	1.5MB/s
	RAMディスク速度	20MB/s
	システム速度	100MB/s
検索機能	あいまい検索	△
	一括照合語数	1,000語
	同義語拡張	△
	異表記展開	○
	論理条件検索	○
	近傍条件検索	○
文脈条件検索	△	

を掛け、RAMディスク上の凝縮本文をストリングサーチエンジンS-S Eへ読み込むことによって行なわれる。最後に、凝縮本文サーチで更に絞り込まれた文書に対して本文サーチを行なう。

本文サーチは、複合条件処理ユニットCRUとストリングサーチエンジンS-S E、及びマルチディスクコントローラMDCに起動を掛けることによって開始される。マルチディスクコントローラMDCの制御によりマルチディスクユニットMDUを構成する各磁気ディスク装置からそれぞれ独立に本文データが読み出され、マルチプレキシングされて高速にストリングサーチエンジンS-S Eへ入力される。

ここでの照合結果は複合条件処理ユニットCRUへ入力され、検索条件式に指定された近傍条件、文脈条件、及び論理条件がこの順にパイプライン的に判別処理される。最終的にヒットした文書の識別子情報がサーチコントローラSCにより、LANを介して検索要求元へ返送される。マルチディスクユニットMDUには本文及び凝縮本文の他に、書誌や抄録なども格納でき、これらを検索対象にすることも、また検索結果に応じて文書識別子を指定して読み出し、LANを介して要求元へ転送することもできる。

### 5. 実験結果

上述したテキストサーチマシンTSM-1に2万2千件の特許明細書(平均容量7.1KB)をデータベースとして登録し、試行実験を行なった。以下この実験により得られた各処理部の基本性能と検索処理性能について報告する。

マルチディスクユニットMDUの最大読み出し速度は10MB/sであり、RAMディスクユニットRDUの最大読み出し速度は20MB/sである。また、ストリングサーチエンジンS-S Eの最大照合速度は20MB/sである。

サロゲート型多段階検索方式における各ブリスサーチステップの処理速度については、図7に示すように、文字成分表サーチ速

度が200KB/s、これに伴うヒット文書の文書識別子変換処理速度が5万3千件/sという実験結果が、また凝縮本文サーチ速度については図8に示すように8.5MB/s、本文サーチ速度については図9に示すように約1MB/sが得られた。

したがって、図10に示すように、文字成分表サーチの絞り込み率を10%、凝縮本文の絞り込み率をその4%と仮定すると、等価的に100MB/sのシステム検索速度が得られる。すなわち、約2万2千件の特許明細書が約1.6秒でフルテキストサーチできることになる。

### 6. おわりに

インデックス情報を用いない自由な言葉による検

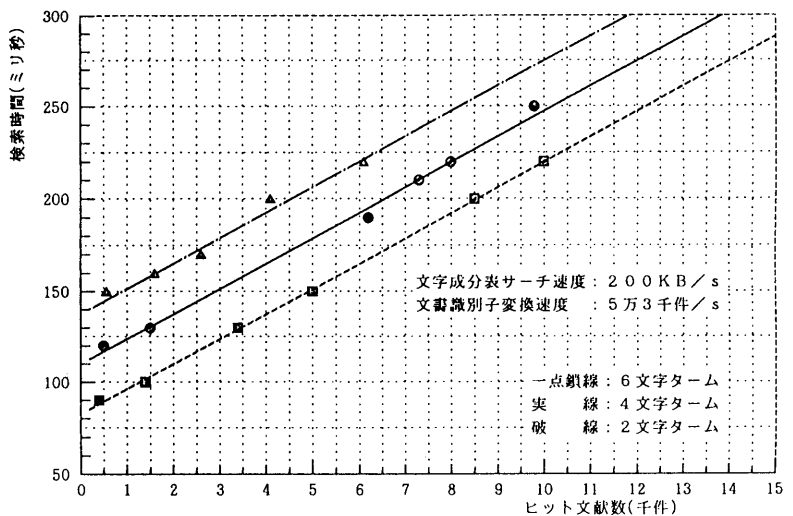


図7. 文字成分表検索処理時間

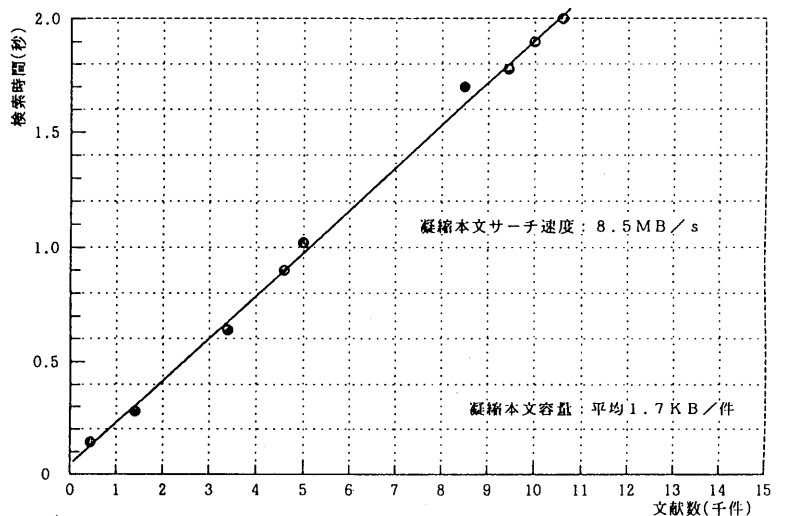


図8. 凝縮本文検索処理時間

索を可能とするスキャン型のフルテキストサーチ方式を開発すると共に、そのプロトタイプ TSM-I を試作し、その実現性を確認した。すなわち、①同義語・異表記検索方式、②サロゲート型検索加速方式、③集合型磁気ディスク並列駆動制御式、④オートマトン型多重文字列照合方式、⑤パイプライン型複合条件判別方式を開発し、テキストサーチマシン TSM-I にインプリメントし、絞り込み率が 0.4% のとき等価的に 100MB/s のシステム検索速度が得られることを確認した。これによれば、500MB (容量 20KB の文書の場合 2万5千件に相当) のテキストデータベースが、実用上十分と考えられる約 5秒でフルテキストサーチできることになる。

今後は、このテキストサーチマシン TSM を大規模なテキストデータベースに適用する際の課題について検討を深めていくと共に、各種テキストデータベースについてフルテキストサーチの有効性を検証して行きたいと考えている。

#### 参考文献

- [1] G.Salton and M.J.McGill, "Introduction to Modern Information Retrieval," McGraw-Hill, New York, 1983.
- [2] G.Salton, "Automatic Text Analysis," Science, Vol. 168, No. 3929, Apr. 1970, pp. 335-343.
- [3] D.C.Blair and M.E.Maron, "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System," Communications of the ACM, Vol. 28, No. 3, 1985, pp. 289-299.
- [4] 諸橋, 「自動索引付け研究の動向」, 情報処理, Vol. 25, No. 9, pp. 918-925 (1984.9).
- [5] L.A.Hollar, "Text Retrieval Computers," Computer, Vol. 12, No. 3, Mar. 1979, pp. 40-50.
- [6] A.V.Aho and M.J.Corasick, "Efficient String Matching," Communications of ACM, Vol. 18, No. 6, June 1975, pp. 333-340.
- [7] R.S.Boyer and J.S.Moore, "A Fast String-Searching Algorithm," Communications of ACM, Vol. 20, No. 10, Aug. 1977, pp. 762-772.
- [8] 若林, 「高速テキスト検索のための専用ハードウェア」, 信学技報, AL85-48, pp. 21-29 (1985).
- [9] 小倉, 山田, 「連想メモリ」, 情報処理, Vol. 27, No. 6, pp. 593-600 (1986.6).
- [10] 高橋, 永井, 山田, 平田, 「ストリング・マッチングハードウェアのアーキテクチャ」, 信学技報, CPSY86-57, pp. 57-68 (1987).

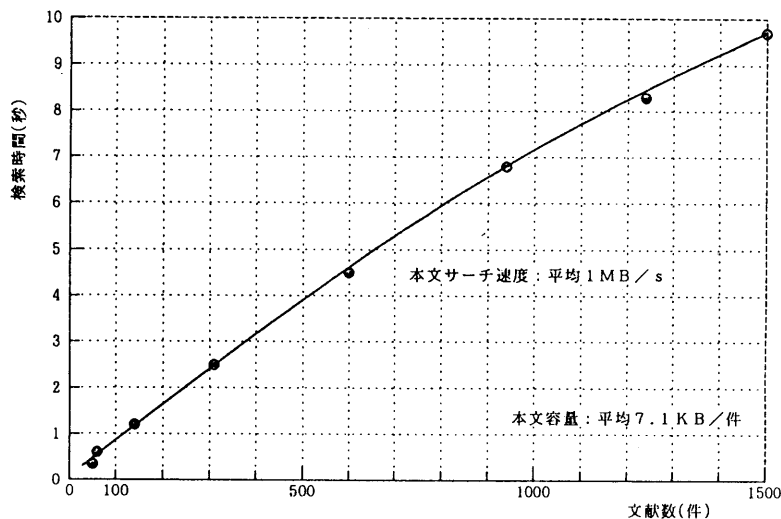


図 9 . 本文検索処理時間

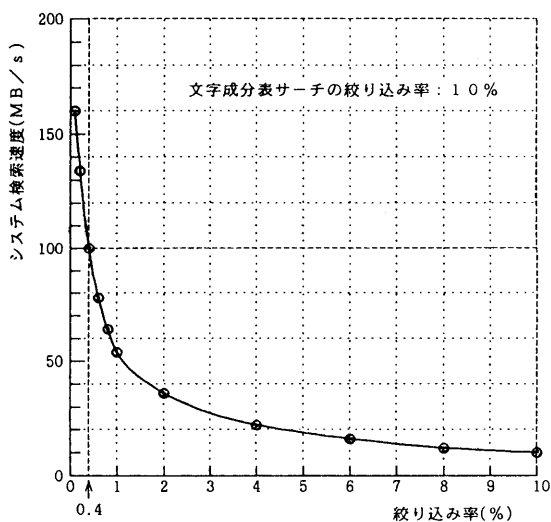


図 10 . システム検索速度

- [11] H.Fujisawa, A.Hatakeyama, and J.Higashino, "A Personal Universal Filing System Based on the Concept-Relation Model," Proc. First Int. Conf. Expert Database Systems, 1986, pp.31-44.
- [12] 藤縄 他, 「知的ファイリングシステムの開発(その1); システムの設計思想と実現方法」, 第33回情報処全大, 4Y-8(1986.10).
- [13] 畠山 他, 「知的ファイリングシステムの開発(その2); 自由語検索における異表記, 異表現解消法」, 第33回情報処全大, 4Y-9(1986.10).
- [14] 川口 他, 「高速文字列検索方式」, 昭62信学全 54(1987).