(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2008-102765
(P2008-102765A)
(43)公開日　平成20年5月1日(2008.5.1)

(51)Int.Cl.		F I		テーマコード（参考）
G06F 17/30	(2006.01)	G06F 17/30	230Z	5B075

審査請求 未請求　請求項の数 5　ＯＬ　（全 22 頁）

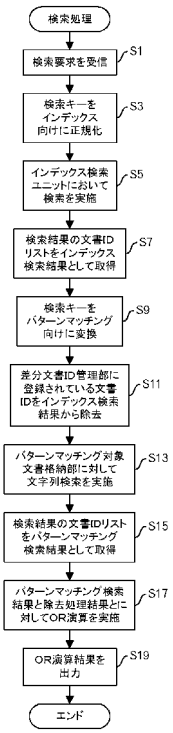
(21)出願番号　特願2006-285077 (P2006-285077)
(22)出願日　平成18年10月19日 (2006.10.19)

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100103528
弁理士　原田　一男
(72)発明者　難波　功
神奈川県川崎市中原区上小田中４丁目１番１号　富士通株式会社内

Ｆターム(参考)　5B075 QS05　UU06

(54)【発明の名称】検索処理方法及び検索システム

(57)【要約】
【課題】大規模な文書集合に対して高速な検索性能を維持しつつ更新遅延を生じさせない。
【解決手段】本検索方法は、文書更新前の検索インデックスを用いて検索を行うインデックス検索ユニットに対して、検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得する工程と、文書更新後の文書内容を保持する文字列検索ユニットに対して、検索要求に係る文字列検索を文書更新後の文書内容に対して実施させ、該当文書の文書ＩＤの第２のリストを取得する工程と、第１のリストと第２のリストと上記文書更新に係る文書の文書ＩＤの第３のリストとを用いて、検索要求に対する検索結果を生成する工程とを含む。このように更新に係る部分の除いた部分についてはインデックス検索ユニットを用いて、更新に係る部分については文字列検索ユニットを用いて検索速度を高速に維持している。
【選択図】図２



【特許請求の範囲】
【請求項１】
文書更新前の検索インデックスを用いて検索を行うインデックス検索ユニットに対して、検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得する第１取得ステップと、
文書更新後の文書内容を保持する文字列検索ユニットに対して、前記検索要求に係る文字列検索を前記文書更新後の文書内容に対して実施させ、該当文書の文書ＩＤの第２のリストを取得する第２取得ステップと、
前記第１のリストと前記第２のリストと文書更新に係る文書の文書ＩＤの第３のリストとを用いて、前記検索要求に対する検索結果を生成する生成ステップと、
を含み、コンピュータにより実行される検索処理方法。
【請求項２】
前記生成ステップが、
前記第１のリストから前記第３のリストに含まれる文書ＩＤを除去し、第４のリストを生成するステップと、
前記第４のリストと前記第２のリストとの和集合を生成するステップと、
を含む請求項１記載の検索処理方法。
【請求項３】
検索要求に応答してインデックス検索を実施するために用いられる第１の検索インデックスと、
更新文書を反映するために用いられる第２の検索インデックスと、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書のデータを格納する第１の更新候補文書格納部と、
前記第２の検索インデックスに対する反映処理に用いられる前記更新文書のデータを格納する第２の更新候補文書格納部と、
検索時に用いられ且つ前記更新文書の文書ＩＤを格納する第１の差分文書ＩＤ管理部と、
検索時に文字列検索を実施するために前記更新文書のデータを格納する第１のパターンマッチング対象文書格納部と、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書の文書ＩＤを格納する第２の差分文書ＩＤ管理部と、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書のデータを格納する第２のパターンマッチング対象文書格納部と、
を有し、
前記第２の検索インデックスに対する前記更新文書の反映処理が完了した場合、
前記反映処理後の前記第２の検索インデックスを、前記第１の検索インデックスとして切り替えて用い、
前記第１の更新候補文書格納部に格納されている前記更新文書のデータを前記第２の更新候補文書格納部に上書き移動させ、
前記第２の差分文書ＩＤ管理部に格納されている前記更新文書の文書ＩＤを前記第１の差分文書ＩＤ管理部に上書き移動させ、
前記第２のパターンマッチング対象文書格納部に格納されている前記更新文書のデータを前記第１のパターンマッチング対象文書格納部に上書き移動させる
検索システム。
【請求項４】
前記第２の検索インデックスに対する前記更新文書の反映処理が開始される前、
前記第１及び第２の更新候補文書格納部に新たな更新要求に係る更新文書のデータを格納し、
前記第１の差分文書ＩＤ管理部に前記新たな更新要求に係る更新文書の文書ＩＤを格納し、

前記第１のパターンマッチング対象文書格納部に前記新たな更新要求に係る更新文書のデータを格納する
請求項３記載の検索システム。
【請求項５】
前記第２の検索インデックスに対する前記更新文書の反映処理が開始された後、
前記第１の更新候補文書格納部をクリアし、
前記第１の更新候補文書格納部に新たな更新要求に係る更新文書のデータを格納し、
前記第１及び第２の差分文書ＩＤ管理部に前記新たな更新要求に係る更新文書の文書ＩＤを格納し、
前記第１及び第２のパターンマッチング対象文書格納部に前記新たな更新要求に係る更新文書のデータを格納する
請求項４記載の検索システム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、文書検索技術に関する。
【背景技術】
【０００２】
近年、更新頻度が高い大規模な文書集合に高速なテキスト検索を行うという要求が生じている。例えば、インターネットやイントラネット上で公開されているブログやニュースは頻繁に更新されているが、これらに対して時間的な遅れなく検索サービスで提供したいという要求がある。また、コールセンタなどにおいて、蓄積された応対記録や直近に受け付けた膨大な応対記録に対して内容の変更に遅延を生ずることなく高速に検索を行いたいという要求もある。
【０００３】
これに対して、文書検索に対して広く実用化されている方式にはインデックスを作成する方式と、検索対象文書が検索キーを含んでいるか否かを判断するため検索対象文書のテキストを走査する文字列検索方式がある。
【０００４】
インデックス方式は、検索対象に対してインデックスを作成し、それにより検索を高速化する方式であり、代表的なものに"Inverted file"がある。この"Inverted file"は、文書中に出現する単語とその文書番号列を保持する索引構造である。"Inverted file"の検索は高速であり、１ＣＰＵ（Central Processing Unit）で１秒あたり数ＧＢから数十ＧＢの検索性能（例えば駆動周波数３ＧＨｚのＣＰＵ）を有し、大規模な文書検索には適している。しかしながら、長大となる文書番号列を圧縮する実装が普通であるため、更新処理は困難であることが多い。
【０００５】
一方、文字列検索（パターンマッチングとも呼ぶ）はインデックスを作成せずに検索対象となるテキストに対して検索目的のパターンが存在するか否かを走査する方式である。文字列検索を用いて検索機構を実装する場合には、検索性能は１ＣＰＵで１秒あたり１０乃至１００ＭＢ程度（例えば駆動周波数３ＧＨｚのＣＰＵ）と低速である。これに対してインデックスを用いないため更新処理は検索対象の更新のみで済み、容易且つ高速である。
【０００６】
なお、特開平８－２７２８０６号公報には、複数の検索方法の実行が可能なデータベース検索システムにおいて、検索者が入力した検索式の条件を自動判断し、各々の検索方法の長所を利用して効率的な検索を行う技術が開示されている。具体的には、データベース格納部内のデータベースを検索するための検索手段として、例えば、インデックス検索手段と全文検索手段を設ける。そして、検索式を入力する入力手段と、検索式を単項式に分割する分割手段と、分割手段で得られた単項式をそれぞれインデックス検索手段と全文検

索手段のいずれかに検索キーとして割り当てる割り当て手段と、インデックス検索手段と全文検索手段からの検索結果について上述の検索式に基づいて論理演算しこの検索式に対応する検索結果として表示手段に出力する演算手段とを設ける。但し、データベースの更新について検討したものではない。

【０００７】

また、特開平９－３１９７６９号公報には、高速なデータ検索を可能とし、システム資源を有効に活用することができる技術が開示されている。具体的には、データベースシステムは、結合された２つのデータ列をそれぞれ備える単一結合データベースと、それらの検索をそれぞれ実行するデータベースドライバと、データベースドライバにおける検索結果を結合するためのインタフェースドライバとを備える。インターフェースドライバは、所望の検索項目を備える単一結合データベースに対応したデータベースドライバに検索を指示し、それらの検索結果を結合することにより、所望の検索結果を取得する。本公報においても、データベースの更新については検討されていない。

【０００８】

さらに、特開平１－９８０２０号公報には、全てのキー値を持つベース索引と共に、キヤツシユ索引が併用され、更新手段を介してキヤツシユ索引に更新情報が一時に蓄積される。検索要求により両索引をマージした検索が行われ、キヤツシユ索引中にベース索引に対応する部分が存在すると、対応するキヤツシユ索引の部分を削除し、ベース索引に追加、削除する技術が開示されている。但し、データベースの更新については検討されていない。

【特許文献１】特開平８－２７２８０６号公報
【特許文献２】特開平９－３１９７６９号公報
【特許文献３】特開平１－９８０２０号公報

【発明の開示】

【発明が解決しようとする課題】

【０００９】

上で述べた"Inverted file"は、インデックス中の文書番号列に圧縮を掛けるために一般に検索サービス中のインデックス更新は困難である。この問題に対処するためにインデックスを多重に保持し、１つを検索用、もう１つをバックグラウンドでの更新用とし用い、バックグラウンドでの更新が終了したらインデックスを切り替えるという方式が採用されている。この方式では更新対象の文書規模にもよるが、原文書の更新に対してインデックスの更新完了時間に数十分程度の遅れが生じる。

【００１０】

また文字列検索では、インデックスの更新は不要であるため更新は高速（具体的には、遅延時間は事実上０とみなせる）であるが、検索速度が低速であるため、数ＧＢ以上の文書を検索しようとすると、検索対象文書を分割して保持するため多数のハードウエア資源が必要となるという問題がある。

【００１１】

従って、本発明の目的は、大規模な文書集合に対して高速な検索性能を維持しつつ更新遅延を生じさせないようにするための技術を提供することである。

【００１２】

また、本発明の他の目的は、ハードウエアコストの増大を抑えつつ文書集合の更新遅延を生じさせないようにするための技術を提供することである。

【課題を解決するための手段】

【００１３】

本発明の第１の態様に係る検索処理方法は、文書更新前の検索インデックスを用いて検索を行うインデックス検索ユニットに対して、検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得する第１取得ステップと、文書更新後の文書内容を保持する文字列検索ユニットに対して、検索要求に係る文字列検索を文書更新後の文書内容に対して実施させ、該当文書の文書ＩＤの第２のリストを取得する第２取得ステップと、第

１のリストと第２のリストと上記文書更新に係る文書の文書ＩＤの第３のリストとを用いて、検索要求に対する検索結果を生成する生成ステップとを含む。

【００１４】

このように更新に係る部分を除いた部分についてはインデックス検索ユニットを用いて、更新に係る部分については文字列検索ユニットを用いて検索速度を高速に維持している。なお、第１取得ステップと第２取得ステップとは並列に実施できる。

【００１５】

また、上で述べた生成ステップが、第１のリストから第３のリストに含まれる文書ＩＤを除去し、第４のリストを生成するステップと、第４のリストと第２のリストとの和集合を生成するステップとを含むようにしてもよい。このようにすれば簡単な論理演算にて正確な検索結果を返すことができる。

【００１６】

また、上で述べた第１取得ステップが、検索要求に係る検索語に対してインデックス検索のための正規化処理を実施するステップを含むようにしてもよい。また、上で述べた第２取得ステップが、検索要求に係る検索語に対して文字列検索のための正規化処理を実施するステップを含むようにしてもよい。検索高速化のためである。

【００１７】

また、本発明の第２の態様に係る検索システムは、検索要求に応答してインデックス検索を実施するために用いられる第１の検索インデックスと、検索時に用いられ且つ更新文書の文書ＩＤを格納する第１の差分文書ＩＤ管理部と、検索時に文字列検索を実施するために更新文書のデータを格納する第１のパターンマッチング対象文書格納部とを有する。そして、第１の検索インデックスを用いて検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得し、検索要求に係る文字列検索を第１のパターンマッチング対象文書格納部に対して実施し、該当文書の文書ＩＤの第２のリストを取得し、第１のリストと第２のリストと第１の差分文書ＩＤ管理部に格納されている文書ＩＤとを用いて、検索要求に対する検索結果を生成するものである。更新遅延を低減しつつ、高速な検索を可能とする構成である。

【００１８】

さらに、本発明の第３の態様に係る検索システムは、検索要求に応答してインデックス検索を実施するために用いられる第１の検索インデックスと、更新文書を反映するために用いられる第２の検索インデックスと、第２の検索インデックスに対する更新文書の反映処理を開始してからの更新文書のデータを格納する第１の更新候補文書格納部と、第２の検索インデックスに対する反映処理に用いられる更新文書のデータを格納する第２の更新候補文書格納部と、検索時に用いられ且つ更新文書の文書ＩＤを格納する第１の差分文書ＩＤ管理部と、検索時に文字列検索を実施するために更新文書のデータを格納する第１のパターンマッチング対象文書格納部と、第２の検索インデックスに対する更新文書の反映処理を開始してからの更新文書の文書ＩＤを格納する第２の差分文書ＩＤ管理部と、第２の検索インデックスに対する更新文書の反映処理を開始してからの更新文書のデータを格納する第２のパターンマッチング対象文書格納部とを有する。そして、第２の検索インデックスに対する更新文書の反映処理が完了した場合、反映処理後の第２の検索インデックスを、第１の検索インデックスとして切り替えて用い、第１の更新候補文書格納部に格納されている更新文書のデータを第２の更新候補文書格納部に上書き移動させ、第２の差分文書ＩＤ管理部に格納されている更新文書の文書ＩＤを第１の差分文書ＩＤ管理部に上書き移動させ、第２のパターンマッチング対象文書格納部に格納されている更新文書のデータを第１のパターンマッチング対象文書格納部に上書き移動させるものである。

【００１９】

このようにすれば、ハードウエアコストの増加を抑えつつ、文書集合に対する更新遅延を生じさせないようにすることができる。すなわち、検索に用いる資源と文書の更新に応じた準備を行う資源とを分けて有し、一斉に上で述べたような処理を実施することによって、検索者への応答に大きな遅延が発生しないようにしつつ、更新遅延も生じさせないよ

うにすることができる。
【００２０】
また、第２の検索インデックスに対する更新文書の反映処理が開始される前には、第１及び第２の更新候補文書格納部に新たな更新要求に係る更新文書のデータを格納するようにしてもよい。その場合、第１の差分文書ＩＤ管理部に新たな更新要求に係る更新文書の文書ＩＤを格納し、第１のパターンマッチング対象文書格納部に新たな更新要求に係る更新文書のデータを格納するようにしてもよい。
【００２１】
さらに、第２の検索インデックスに対する更新文書の反映処理が開始された後には、第１の更新候補文書格納部をクリアするようにしてもよい。その場合、第１の更新候補文書格納部に新たな更新要求に係る更新文書のデータを格納し、第１及び第２の差分文書ＩＤ管理部に新たな更新要求に係る更新文書の文書ＩＤを格納し、第１及び第２のパターンマッチング対象文書格納部に新たな更新要求に係る更新文書のデータを格納するようにしてもよい。
【００２２】
さらに、第１の検索インデックスを用いて検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得し、検索要求に係る文字列検索を第１のパターンマッチング対象文書格納部に対して実施し、該当文書の文書ＩＤの第２のリストを取得し、第１のリストと第２のリストと第１の差分文書ＩＤ管理部に格納されている文書ＩＤとを用いて、検索要求に対する検索結果を生成するようにしてもよい。
【００２３】
また、第１及び第２の更新候補文書格納部に格納される更新文書のデータが、検索インデックス向けに正規化されており、第１及び第２のパターンマッチング対象文書格納部に格納される更新文書のデータが、文字列検索向けに正規化されている場合もある。
【００２４】
本発明にかかる検索処理方法をコンピュータに実行させるためのプログラムや検索システムをコンピュータに実施させるためのプログラムを作成することができ、当該プログラムは、例えばフレキシブル・ディスク、ＣＤ－ＲＯＭ、光磁気ディスク、半導体メモリ、ハードディスク等の記憶媒体又は記憶装置に格納される。また、ネットワークを介してディジタル信号にて頒布される場合もある。なお、処理途中のデータについては、コンピュータのメモリ等の記憶装置に一時保管される。
【発明の効果】
【００２５】
本発明によれば、大規模な文書集合に対して高速な検索性能を維持しつつ更新遅延を生じさせないようにすることができる。
【００２６】
また、本発明の他の側面によれば、ハードウエアコストの増大を抑えつつ文書集合の更新遅延を生じさせないようにすることができる。
【発明を実施するための最良の形態】
【００２７】
図１に本発明の一実施の形態に係る検索システムのシステム概要図を示す。本実施の形態における検索システムは、ユーザ等から検索要求などの処理要求を受け付ける処理要求受信部１と、検索要求に係る検索語や文書更新要求に係る文書などに対してインデックス検索ユニット向けの正規化処理及び文字列検索ユニット向けの正規化処理を実施する正規化処理部３と、インデックス検索ユニットＡ及びＢと、正規化処理部３からのデータに応じてインデックス検索ユニットＡ及びＢに対して以下で詳細に述べる制御を実施するインデックス検索コントローラ５と、文字列検索ユニットＡ及びＢと、文字列検索ユニットＡに対して検索要求に係る文字列検索を行うパターンマッチング処理部９と、正規化処理部３からのデータに応じてパターンマッチング処理部９並びに文字列検索ユニットＡ及びＢに対して以下で詳細に述べる制御を実施する文字列検索コントローラ７とを有する。

【００２８】
インデックス検索ユニットＡは、検索用であって、検索インデックスを有するインデックス処理部１１１と、更新候補文書格納部１１３とを有する。インデックス検索ユニットＢは、更新用であって、検索インデックスを有するインデックス処理部１３１と、更新候補文書格納部１３３とを有する。インデックス検索ユニットＡ及びＢは、周知のインデックス検索を実施するためのものであり、インデックス処理部１１１には所定のタイミングで生成された検索インデックスが保持されている。
【００２９】
インデックス検索コントローラ５と文字列コントローラ７とは連係して動作する。
【００３０】
文字列検索ユニットＡは、検索用であって、差分文書ＩＤ管理部１５１と、パターンマッチング対象文書格納部１５３とを有する。文字列検索ユニットＢは、更新用であって、差分文書ＩＤ管理部１７１と、パターンマッチング対象文書格納部１７３とを有する。文字列検索ユニットＡ及びＢは、周知の文字列検索（すなわちパターンマッチング）を実施するためのものであって、所定のタイミング以降に更新された文書内容をパターンマッチング対象文書格納部１５３に保持すると共に、差分文書ＩＤ管理部１５１は所定のタイミング以降に更新された文書の文書ＩＤのリストを保持している。
【００３１】
なお、検索システムは、例えばネットワークに接続されており、ネットワークに接続された他のコンピュータから検索要求などを受信するものとして説明するが、スタンドアローン形式で実施するようにしても良い。
【００３２】
次に、図２を用いて図１に示した検索システムにおける検索処理を説明する。まず、処理要求受信部１は、検索キー（すなわち検索語）を含む検索要求をネットワークを介して受信し（ステップＳ１）、検索キーを含む検索要求を正規化処理部３に出力する。正規化処理部３は、処理要求受信部１から検索キーを含む検索要求を受信すると、検索キーをインデックス向けに正規化し、メインメモリなどの記憶装置に格納する（ステップＳ３）。正規化処理部３の処理内容については、後にまとめて説明するが、共通の正規化規則に従ってインデックス検索と文字列検索の差を抑制するものである。例えば、文字列検索では文書中の任意の文字列部分にマッチするため、ｔｈｉｓを含む文書をｉｓで検索するとヒットするが、インデックスが単語単位である場合にｔｈｉｓの部分列はヒットしない。このため同一の単語を含む文章であってもインデックス側で検索される場合と文字列検索側で検索される場合で結果に差が生じる可能性があるためこれを抑制するものが正規化規則である。本実施の形態では、共通の正規化としてケース・ホールディング（case folding）を実施し文字列検索に対しては、単語を＃で囲むという正規化を行い、インデックスに対しては、これが単語単位で索引を作成すると想定しステミング（stemming）を抑制するとしている。そして、正規化処理部３は、正規化された検索キーを含む検索要求を、インデックス検索コントローラ５に出力する。
【００３３】
インデックス検索コントローラ５は、正規化処理部３から正規化された検索キーを含む検索要求を受信すると、インデックス検索ユニットＡに当該検索要求を転送する。
【００３４】
インデックス検索ユニットＡは、インデックス検索コントローラ５から検索要求を受信すると、そのインデックス処理部１１１により周知のインデックス検索を実施する（ステップＳ５）。そして、インデックス検索ユニットＡのインデックス処理部１１１は、検索結果をインデックス検索コントローラ５に出力する。検索結果は文書ＩＤのリストであって、ここでは例えば文書１、文書３、文書５及び文書１３とする。インデックス検索コントローラ５は、インデックス処理部１１１から検索結果の文書ＩＤリストを受信し（ステップＳ７）、例えばメインメモリなどの記憶装置に格納すると共に、文字列検索コントローラ７に検索結果の文書ＩＤリストを出力する。文字列検索コントローラ７は、インデッ

クス検索の検索結果である文書ＩＤリストを受信する。

【００３５】

また、ステップＳ５及びＳ７に並行して、正規化処理部３は、処理要求受信部１から受信した検索要求に含まれる検索キーをパターンマッチング向け（すなわち文字列検索向け）に正規化（すなわち変換）する（ステップＳ９）。例えば、共通の正規化規則に従ってケース・ホールディング（case folding）を実施し、アルファベット連続のトークン（token）の前後を＃で囲む。例えば、Ｔｈｉｓであれば、＃ｔｈｉｓ＃とする。正規化処理部３は、パターンマッチング向けに変換された検索キーを含む検索要求を文字列検索コントローラ７に出力する。

【００３６】

さらに、文字列検索コントローラ７は、文字列検索ユニットＡの差分文書ＩＤ管理部１５１から、当該差分文書ＩＤ管理部１５１に登録されている文書ＩＤのリストを取得し、インデックス検索コントローラ５から受信したインデックス検索の検索結果である文書ＩＤリストから、取得した文書ＩＤのリストを引き算して除去して、残余の文書ＩＤのリストを例えばメインメモリなどの記憶装置に格納する（ステップＳ１１）。差分文書ＩＤ管理部１５１は、インデックス処理部１１１において用いられている現在の検索インデックスが生成された後に更新された文書の文書ＩＤのリストを保持しているので、たとえインデックス検索の結果に、差分文書ＩＤ管理部１５１が保持している文書ＩＤのリストに含まれる文書ＩＤが含まれていても、それは信用できない。従って、本ステップにおいて、インデックス検索の検索結果である文書ＩＤのリストから、差分文書ＩＤ管理部１５１が保持している文書ＩＤのリストに含まれる文書ＩＤを除去する。例えば、差分文書ＩＤ管理部１５１が保持している文書ＩＤが、文書５、文書２８、文書３４であるとすると、文書５がインデックス検索の検索結果と重複するので除去され、残りは文書１、文書３、文書１３となる。

【００３７】

文字列検索コントローラ７は、正規化処理部３から、変換後の検索キーを含む検索要求を受信すると、パターンマッチング処理部９へ変換後の検索キーを出力して、パターンマッチング処理部９にパターンマッチング対象文書格納部１５３に対する周知の文字列検索を実施させる（ステップＳ１３）。パターンマッチング処理部９は、変換後の検索キーを用いて、検索に用いられた検索インデックス生成後に更新された文書の内容を保持するパターンマッチング対象文書格納部１５３を走査し、該当する文書ＩＤを特定し、該当文書の文書ＩＤのリストを文字列検索コントローラ７に出力する。文字列検索コントローラ７は、パターンマッチング処理部９から、文字列検索の検索結果である文書ＩＤリストをパターンマッチング検索結果として受信し、例えばメインメモリなどの記憶装置に格納する（ステップＳ１５）。例えば、文書１、文書８、文書２８及び文書３４が処理結果として得られたものとする。

【００３８】

そして、文字列検索コントローラ７は、パターンマッチング検索結果の文書ＩＤリストとステップＳ１１の処理結果として得た残余の文書ＩＤのリスト（すなわち除去処理結果）とに対してＯＲ演算を実施（すなわち和集合を生成）し、例えばメインメモリなどの記憶装置に格納する（ステップＳ１７）。これによって、現在の検索インデックス生成後に更新された文書の更新内容を反映した形で検索結果が生成される。例えば、残余の文書ＩＤのリストが、文書１、文書３、文書１３であって、パターンマッチング検索結果が文書１、文書８、文書２８及び文書３４であるから、和集合としては、文書１、文書３、文書８、文書１３、文書２８及び文書３４となる。

【００３９】

その後、例えば、文字列検索コントローラ７は、要求元のユーザ端末に、ＯＲ演算の結果（すなわち文書ＩＤリスト）を出力する（ステップＳ１９）。

【００４０】

このようにすれば、大量の文書についてはインデックス検索ユニットＡによるインデッ

クス検索を実施し、少量の更新文書については文字列検索ユニットＡに対するパターンマッチング処理部９による文字列検索を並行に行うことができ、検索の高速性が保持しつつ、正しい検索結果を得ることができるようになる。

【００４１】

次に、図３乃至図２６を用いて、更新処理について説明する。まず、図３を用いてインデックス差分更新開始処理について説明する。まず、管理者端末からインデックス差分更新開始処理を指示されるか又は定期的又は任意のタイミングで自動的にインデックス差分更新開始処理が指示されると、例えばインデックス検索コントローラ５は、自身のインデックス差分更新開始フラグをセットすると共に、文字列検索コントローラ７におけるインデックス差分更新開始フラグをセットする（ステップＳ２１）。そして、インデックス検索コントローラ７は、インデックス検索ユニットＡの更新候補文書格納部１１３をクリアする（ステップＳ２３）。更新候補文書格納部１１３には、以下で説明するが、現在の検索インデックス生成開始後に更新された文書の内容が格納されており、それがクリアされる。なお、バックアップのために他の記憶領域に格納しておくようにしても良い。

【００４２】

さらに、インデックス検索コントローラ７は、インデックス検索ユニットＢのインデックス処理部１３１に、自身の検索インデックスに対して更新候補文書格納部１３３に保持されている文書を反映するための更新処理を実施させる（ステップＳ２５）。更新候補文書格納部１３３は、インデックス検索ユニットＢの検索インデックスに対する前回の更新処理が開始した後に受信した更新要求に係る文書の内容を保持している。この前回の更新処理が開始した後の更新内容を反映するために、検索インデックスの更新を開始するものである。

【００４３】

次に、他のコンピュータや、検索システムの更新文書取得部（図示せず）、ユーザ端末などから検索対象の文書の更新要求を受信した際の処理について、図４乃至図２２を用いて説明する。まず、他のコンピュータや、検索システムの更新文書取得部、ユーザ端末などから、処理要求受信部１は、更新後の文書内容を含む文書更新要求を受信し（ステップＳ３１）、正規化処理部３に出力する。そして、正規化処理部３は、文書更新要求を受信した旨をインデックス検索コントローラ５及び文字列検索コントローラ７に通知する。

【００４４】

そうすると、インデックス検索コントローラ５は、インデックス検索ユニットＡ及びＢに書き込みロックをかけ、文字列検索コントローラ７は、文字列検索ユニットＡ及びＢに書き込みロックをかける（ステップＳ３３）。

【００４５】

また、正規化処理部３は、文書更新要求に係る文書の内容をインデックス向けに正規化し、例えばメインメモリなどの記憶装置に格納すると共に、インデックス検索コントローラ５に出力する（ステップＳ３５）。文書更新要求に係る正規化された文書内容を受信したインデックス検索コントローラ５は、自身のインデックス差分開始フラグがＯＮであるか判断する（ステップＳ３７）。インデックス差分開始フラグがＯＦＦである場合には、インデックス検索コントローラ５は、インデックス検索ユニットＢの更新候補文書格納部１３３に、正規化された文書更新要求に係る文書の内容を登録する（ステップＳ３９）。

【００４６】

また、インデックス差分開始フラグがＯＮである場合であってもＯＦＦであっても、インデックス検索コントローラ５は、インデックス検索ユニットＡの更新候補文書格納部１１３に、正規化された文書更新要求に係る文書の内容を登録する（ステップＳ４１）。

【００４７】

図５乃至図８を用いて、ここまでの処理の内容について説明する。例えば図５及び図６に示すように、インデックス検索ユニットＡのインデックス処理部１１１の検索インデックスも、インデックス検索ユニットＢのインデックス処理部１３１の検索インデックスも空であり、さらにインデックス検索ユニットＡの更新候補文書格納部１１３もインデック

ス検索ユニットＢの更新候補文書格納部１３３も空である状態を想定する。

【００４８】

そして、文書ＩＤ「００００１」で文書内容「Ｔｈｉｓ　ｉｓ　ａ　ｂｏｏｋ」と、文書ＩＤ「００００２」で文書内容「Ｔｈｉｓ　ｉｓ　ａ　ｐｅｎ」と、文書ＩＤ「００００３」で文書内容「Ｔｈｉｓ　ｗａｓ　ａ　ｂｏｏｋ」と、文書ＩＤ「００００４」で文書内容「Ｔｈａｔ　ｉｓ　ａ　ｂｏｏｋ」と、文書ＩＤ「００００５」で文書内容「Ｔｈａｔ　ｗａｓ　ａ　ｐｅｎ」とが文書更新要求で指定されるとする。

【００４９】

そうすると、インデックス差分更新開始フラグがＯＦＦであれば、ステップＳ３９及びＳ４１で、図７及び図８に示すように、正規化された文書内容（「Ｔｈｉｓ」が「ｔｈｉｓ」に正規化され、「Ｔｈａｔ」が「ｔｈａｔ」に正規化される）が、インデックス検索ユニットＡの更新候補文書格納部１１３と、インデックス検索ユニットＢの更新候補文書格納部１３３とに登録される。なお、インデックス差分更新開始フラグがＯＮであれば、インデックス検索ユニットＡの更新候補文書格納部１１３にのみ更新内容が登録される。

【００５０】

図４の説明に戻って、ステップＳ３７乃至Ｓ４１とは並列に、正規化処理部３は、文書更新要求に係る文書の内容をパターンマッチング向けに変換（すなわち正規化）し、例えばメインメモリなどの記憶装置に格納すると共に、文字列検索コントローラ７に出力する（ステップＳ４３）。処理は端子Ａを介して図９の処理フローに移行する。

【００５１】

文字列検索コントローラ７は、自身のインデックス差分更新開始フラグがＯＮであるか判断する（ステップＳ４４）。インデックス差分更新開始フラグがＯＮである場合には、文字列検索コントローラ７は、文字列検索ユニットＢのパターンマッチング対象文書格納部１７３に、変換後の、文書検索要求に係る文書の内容を登録する（ステップＳ４５）。また、文字列検索コントローラ７は、文字列検索ユニットＢの差分文書ＩＤ管理部１７１に、文書更新要求に係る文書の文書ＩＤを登録する（ステップＳ４７）。

【００５２】

また、インデックス差分更新開始フラグがＯＦＦの場合であってもＯＮの場合であっても、文字列検索コントローラ７は、文字列検索ユニットＡのパターンマッチング対象文書格納部１５１に、変換後の、文書検索要求に係る文書の内容を登録する（ステップＳ４９）。また、文字列検索コントローラ７は、文字列検索ユニットＡの差分文書ＩＤ管理部１５１に、文書更新要求に係る文書の文書ＩＤを登録する（ステップＳ５１）。

【００５３】

図１０乃至図１３を用いてステップＳ４４乃至Ｓ５１の処理について説明する。まず、前提として、図１０及び図１１に示すように、文字列検索ユニットＡの差分文書ＩＤ管理部１５１も、文字列検索ユニットＢの差分文書ＩＤ管理部１７１も空であり、さらに文字列検索ユニットＡのパターンマッチング対象文書格納部１５３も文字列検索ユニットＢのパターンマッチング対象文書格納部１７３も空であるとする。

【００５４】

その後上で述べたような文書更新要求を受信すると、図１２及び図１３に示すように、正規化された文書内容（「Ｔｈｉｓ　ｉｓ　ａ　ｂｏｏｋ」が「＃ｔｈｉｓ＃　＃ｉｓ＃　＃ａ＃　＃ｂｏｏｋ＃」といったように正規化される）が、インデックス差分更新開始フラグがＯＦＦであれば、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３のみに格納される。また、更新文書の文書ＩＤ（「００００１」乃至「００００５」）も、文字列検索ユニットＡの差分文書ＩＤ管理部１５１のみに登録される。

【００５５】

一方、インデックス差分更新開始フラグがＯＮであれば、正規化された文書内容は、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３にも文字列検索ユニットＢのパターンマッチング対象文書格納部１７３にも格納される。また、更新文書の文書ＩＤも、文字列検索ユニットＡの差分文書ＩＤ管理部１５１にも文字列検索ユニットＢの差

分文書ＩＤ管理部１７１にも登録される。

【００５６】

そして、インデックス検索コントローラ５は、インデックス検索ユニットＡ及びＢの書き込みロックを解除し、文字列検索コントローラ７は、文字列検索ユニットＡ及びＢの書き込みロックを解除する（ステップＳ５３）。

【００５７】

このようにして高速検索を可能にしつつ文書更新に対処する。

【００５８】

なお、本実施の形態を分かりやすくするために、図５乃至図８及び図１０乃至図１３の後に、以下のような処理が行われた場合について具体的に説明する。まず、「ＩＳ」を検索キーとする検索要求を受信した場合には、正規化処理部３はインデックス検索向けに「ｉｓ」に正規化し、文字列検索向けに「＃ｉｓ＃」に変換する。この「ｉｓ」をベースにインデックス検索ユニットＡのインデックス処理部１１１によりインデックス検索が行われるが、検索インデックスが空であるので、インデックス検索ユニットＡのインデックス処理部１１１からの検索結果は「空」である。また、文字列検索コントローラ７が文字列検索ユニットＡの差分文書ＩＤ管理部１５１から保持する文書ＩＤのリストを受け取っても、インデックス検索コントローラ５から受信したインデックス検索の検索結果が「空」であるから、引き算した結果も「空」となる。そこで、パターンマッチング処理部９が、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に対して「＃ｉｓ＃」をベースにパターンマッチング処理を実施すると、文書ＩＤ「００００１」「００００２」「００００４」を得ることができる。そして、文字列検索コントローラ７は、「空」と文書ＩＤ「００００１」「００００２」「００００４」との和集合を生成し、検索結果として文書ＩＤ「００００１」「００００２」「００００４」を出力する。

【００５９】

次に、インデックス差分更新開始処理が実施されると、図３の処理に従って処理が行われ、図１４に示されるように、インデックス検索ユニットＡの更新候補文書格納部１１３はクリアされて、「空」となる。また、図１５に示されるように、インデックス検索ユニットＢのインデックス処理部１３１は、更新候補文書格納部１３３に格納されている更新文書の内容を自身の検索インデックスに反映させる処理（すなわち更新処理）を開始する。

【００６０】

その後、文書ＩＤ「００００１」の文書内容を「Ｔｈｏｓｅ　ａｒｅ　ｂｏｏｋｓ」に変更する文書更新要求を受信した場合、正規化処理部３は、インデックス向けに「ｔｈｏｓｅ　ａｒｅ　ｂｏｏｋｓ」と正規化し、図１６に示すように、インデックス検索コントローラ５は、インデックス検索ユニットＡの更新候補文書格納部１１３に文書ＩＤと更新後の文書内容を登録する。但し、図１７に示すように、インデックス検索ユニットＢのインデックス処理部１３１は、更新処理を実施中であり、更新候補文書格納部１３３には更新を行わない。

【００６１】

また、正規化処理部３は、パターンマッチング向けに「＃ｔｈｏｓｅ＃　＃ａｒｅ＃　＃ｂｏｏｋｓ＃」に変換し、文字列検索コントローラ７は、文字列検索ユニットＡの差分文書ＩＤ管理部１５１に、文書更新要求に係る文書の文書ＩＤ「００００１」を登録する。但し、既に文書ＩＤ「００００１」は登録済みであるので、変化はない。また、図１８に示すように、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に、文書更新要求に係る文書の文書ＩＤ及び変換後の更新内容を登録する。図１８では、更新部分に下線を引いてある。さらに、図１９に示すように、文字列検索ユニットＢのパターンマッチング対象文書格納部１７３に、文書更新要求に係る文書の文書ＩＤ及び変換後の更新内容を登録する。図１３で示したように、インデックス差分更新開始フラグがＯＦＦである場合には、差分文書ＩＤ管理部１７１にもパターンマッチング対象文書格納部１７３にもデータは登録されないので、今回の文書更新要求に係る文書の文書ＩＤ及び変換後の

更新内容が初めて登録されることになる。

【００６２】

さらに、文書ＩＤ「００００７」で文書内容「Ｔｈａｔ　ｉｓ　ａ　ｐｅｎｃｉｌ」という文書を追加する文書更新要求を受信した場合、正規化処理部３は、インデックス向けに「ｔｈａｔ　ｉｓ　ａ　ｐｅｎｃｉｌ」と正規化し、図２０に示すように、インデックス検索コントローラ５は、インデックス検索ユニットＡの更新候補文書格納部１１３に文書ＩＤと追加の文書内容を登録する。但し、インデックス検索ユニットＢのインデックス処理部１３１は、更新処理を実施中であり、更新候補文書格納部１３３には更新を行わない。

【００６３】

また、正規化処理部３は、パターンマッチング向けに「＃ｔｈａｔ＃　＃ｉｓ＃　＃ａ＃　＃ｐｅｎｃｉｌ＃」に変換し、図２１に示すように、文字列検索コントローラ７は、文字列検索ユニットＡの差分文書ＩＤ管理部１５１に、文書更新要求に係る文書の文書ＩＤ「００００７」を登録する。また、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に、文書更新要求に係る文書の文書ＩＤ及び追加の文書内容を登録する。図２１では、更新部分に下線を引いてある。さらに、図２２に示すように、文字列検索ユニットＢのパターンマッチング対象文書格納部１７３に、文書更新要求に係る文書の文書ＩＤ及び追加の文書内容を登録する。

【００６４】

ここで「ＩＳ」を検索キーとする検索要求を受信した場合には、正規化処理部３はインデックス検索向けに「ｉｓ」に正規化し、文字列検索向けに「＃ｉｓ＃」に変換する。この「ｉｓ」をベースにインデックス検索ユニットＡのインデックス処理部１１１によりインデックス検索が行われるが、検索インデックスが空であるので、インデックス検索ユニットＡのインデックス処理部１１１からの検索結果は「空」である。

【００６５】

また、文字列検索コントローラ７が文字列検索ユニットＡの差分文書ＩＤ管理部１５１から保持する文書ＩＤのリストを受け取っても、インデックス検索コントローラ５から受信したインデックス検索の検索結果が「空」であるから、引き算した結果も「空」となる。そこで、パターンマッチング処理部９が、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に対して変換化後の「＃ｉｓ＃」をベースにパターンマッチング処理を実施すると、文書ＩＤ「００００２」「００００４」「００００７」を得ることができる。そして、文字列検索コントローラ７は、「空」と文書ＩＤ「００００２」「００００４」「００００７」との和集合を生成し、検索結果として文書ＩＤ「００００２」「００００４」「００００７」を出力する。

【００６６】

次に、図２３乃至図２５を用いて、インデックス検索ユニットＢのインデックス処理部１３１による更新文書の反映が終了した場合に実施される差分更新終了処理を説明する。インデックス検索ユニットＢのインデックス処理部１３１は、更新文書の反映処理を終了すると、インデックス検索コントローラ５に当該終了を通知する。インデックス検索コントローラ５は、文字列検索コントローラ７にも終了を通知する。そうすると、インデックス検索コントローラ５は、インデックス検索ユニットＡ及びＢに対する書き込みロックを行い、さらに文字列検索コントローラ７は、文字列検索ユニットＡ及びＢに対する書き込みロックを実施する（ステップＳ６１）。

【００６７】

そして、インデックス検索コントローラ５は、インデックス検索ユニットＢの検索インデックスをインデックス検索ユニットＡのインデックス処理部１１１にコピーする（ステップＳ６３）。また、インデックス検索コントローラ５は、インデックス検索ユニットＡの更新候補文書格納部１１３の格納データを、インデックス検索ユニットＢの更新候補文書格納部１３３に上書きコピーする（ステップＳ６５）。そして、インデックス検索コントローラ５は、インデックス検索ユニットＡの更新候補文書格納部１１３をクリアする（

ステップＳ６７）。

【００６８】

更新文書の反映処理を実施している間、インデックス検索ユニットＢの更新候補文書格納部１３３は更新されず、インデックス検索ユニットＡの更新候補文書格納部１１３に更新文書の内容が蓄積されるので、検索インデックスに対する更新文書の反映処理が完了すれば、上で述べたような処理を実施し、次に行われる更新文書の反映処理に備える。

【００６９】

さらに、文字列検索コントローラ７は、文字列検索ユニットＢの差分文書ＩＤ管理部１７１の格納データを、文字列検索ユニットＡの差分文書ＩＤ管理部１５１に上書きコピーを行う（ステップＳ６９）。文字列検索コントローラ７は、文字列検索ユニットＢのパターンマッチング対象文書格納部１７３の格納データを、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に上書きコピーする（ステップＳ７１）。また、文字列検索コントローラ７は、文字列検索ユニットＢの差分文書ＩＤ管理部１７１及びパターンマッチング対象文書格納部１７３をクリアする（ステップＳ７３）。

【００７０】

検索インデックスに対する更新文書の反映処理を実施している間、文字列検索ユニットＡも文字列検索ユニットＢも更新されるが、更新文書の反映処理が実施されていない間は文字列検索ユニットＡのみが更新されるので、文字列検索ユニットＢには更新文書の反映処理を実施している間のみの更新文書の内容が蓄積される。従って、検索インデックスが更新された後には、文字列検索ユニットＢの内容で文字列検索ユニットＡを置換することで、正確な検索を行うことができるようになる。

【００７１】

また、インデックス検索コントローラ５は、インデックス検索ユニットＡ及びＢに対する書き込みロックを解除し、文字列検索ユニットＡ及びＢに対する書き込みロックを解除する（ステップＳ７５）。

【００７２】

上で述べた具体例を用いて差分更新終了処理を説明する。図２４に示すように、インデックス検索ユニットＢのインデックス処理部１３１により生成された検索インデックス（文書ＩＤ「００００１」乃至「００００５」に対するインデックス）を、インデックス検索ユニットＡのインデックス処理部１１３にコピーする。また、インデックス検索ユニットＡの更新候補文書格納部１１３に格納された文書ＩＤ「００００１」及び「００００７」及びその更新内容を、インデックス検索ユニットＢの更新候補文書格納部１３３に上書き移動する。これによって、インデックス検索ユニットＡの更新候補文書格納部１１３は空になる。

【００７３】

さらに、文字列検索ユニットＢの差分文書ＩＤ管理部１７１の格納データである文書ＩＤ「００００１」及び「００００７」を、文字列検索ユニットＡの差分文書ＩＤ管理部１５１に上書き移動する。同様に、文字列検索ユニットＢのパターンマッチング対象文書格納部１７３の格納データである文書ＩＤ「００００１」及び「００００７」並びにその内容を、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に上書き移動する。これによって、文字列検索ユニットＢの差分文書ＩＤ管理部１７１及びパターンマッチング対象文書格納部１７３は空になる。

【００７４】

図２４及び図２５に示された状態になった後に、「ＩＳ」を検索キーとする検索要求を受信した場合には、正規化処理部３はインデックス検索向けに「ｉｓ」に正規化し、文字列検索向けに「＃ｉｓ＃」に変換する。この「ｉｓ」をベースにインデックス検索ユニットＡのインデックス処理部１１１によりインデックス検索が行われるが、図１７の更新候補文書格納部１３３に格納されているデータをベースに検索インデックスが構成されているので、文書ＩＤ「００００１」「００００２」「００００４」がインデックス検索の検索結果として抽出される。また、文字列検索コントローラ７が文字列検索ユニットＡの差

分文書ＩＤ管理部１５１から保持する文書ＩＤのリスト（文書ＩＤ「００００１」及び「００００７」）を受け取って、インデックス検索コントローラ５から受信したインデックス検索の検索結果から引き算すると、文書ＩＤ「００００２」「００００４」が得られる。

【００７５】

また、パターンマッチング処理部９が、文字列検索ユニットＡのパターンマッチング対象文書格納部１５３に対して「＃ｉｓ＃」をベースにパターンマッチング処理を実施すると、文書ＩＤ「００００７」を得ることができる。そして、文字列検索コントローラ７は、文書ＩＤ「００００７」と文書ＩＤ「００００２」「００００４」との和集合を生成すると、検索結果として文書ＩＤ「００００２」「００００４」「００００７」を出力する。

【００７６】

このように検索インデックスによって大量の文書に対する検索速度を維持すると共に、少量の更新文書に対しては並行して実施される文字列検索によって対処して、更新文書の文書ＩＤと併せて適切に合成することによって、正しい検索結果を高速に得ることができるようになる。

【００７７】

次に、図２６を用いて正規化処理部３の処理内容をまとめておく。まず、インデックス及び文字列検索に共通の正規化処理を実施し、処理結果を例えばメインメモリなどの記憶装置に格納する（ステップＳ８１）。例えば、「Ｔｈｅｓｅ　ａｒｅ　ｂｏｏｋｓ」であれば、「ｔｈｅｓｅ　ａｒｅ　ｂｏｏｋｓ」とcase folding変換される。そして、文字列検索向けの正規化処理であるか判断し（ステップＳ８３）、インデックス検索向けの正規化処理であれば、インデックス検索固有のマッチング処理を抑制する正規化を実施する（ステップＳ８７）。この結果は例えば、「ｔｈｅｓｅ　ａｒｅ　ｂｏｏｋｓ」となる。本実施の形態では抑制する処理は英語の情報検索でよく用いられるステミング（stemming）を想定しており、この処理は適用されないので、「ｂｏｏｋｓ」は「ｂｏｏｋ」に変換されない。

【００７８】

一方、文字列検索向けの正規化処理の場合には、文字列検索固有のマッチング処理を抑制する正規化処理を実施する（ステップＳ８５）。文字列検索固有のマッチングとしてはその性質上任意の文字列にマッチするというものがあるが、本実施の形態ではインデックスが単語単位であり基本的に空白区切りで単語を認定していることを想定しているためこれにあわせるため空白で囲まれている文字列の前後に＃を付与する処理を行っている。これにより入力は「＃ｔｈｅｓｅ＃　＃ａｒｅ＃　＃ｂｏｏｋｓ＃」というように変換される。正規化処理の選択は対象となっている言語でどのようにインデックスと文字列検索で検索結果の整合性をとるかに依存するため本実施の形態以外にさまざまな方式が可能である。

【００７９】

このようにすれば、高速な検索を実施することができるようになる。

【００８０】

以上本発明の一実施の形態を説明したが、本発明はこれに限定されるものではない。例えば、図１に示した機能ブロック図は一例であって、必ずしも実際のプログラムモジュールとは対応しない。

【００８１】

さらに、処理フローでは一列に並んでいるステップであっても処理結果が変わらない限り並列実行できるものもある。また、処理結果が変わらない限り順番を入れ替えることも可能である。

【００８２】

なお、本実施の形態における検索システムは、コンピュータ装置であって、図２７に示すように、メモリ２５０１とＣＰＵ２５０３とハードディスク・ドライブ（ＨＤＤ）２５

０５と表示装置２５０９に接続される表示制御部２５０７とリムーバブル・ディスク２５１１用のドライブ装置２５１３と入力装置２５１５とネットワークに接続するための通信制御部２５１７とがバス２５１９で接続されている。オペレーティング・システム（ＯＳ：Operating System）及び本実施例における処理を実施するためのアプリケーション・プログラムは、ＨＤＤ２５０５に格納されており、ＣＰＵ２５０３により実行される際にはＨＤＤ２５０５からメモリ２５０１に読み出される。必要に応じてＣＰＵ２５０３は、表示制御部２５０７、通信制御部２５１７、ドライブ装置２５１３を制御して、必要な動作を行わせる。また、処理途中のデータについては、メモリ２５０１に格納され、必要があればＨＤＤ２５０５に格納される。本発明の実施例では、上で述べた処理を実施するためのアプリケーション・プログラムはリムーバブル・ディスク２５１１に格納されて頒布され、ドライブ装置２５１３からＨＤＤ２５０５にインストールされる。インターネットなどのネットワーク及び通信制御部２５１７を経由して、ＨＤＤ２５０５にインストールされる場合もある。このようなコンピュータ装置は、上で述べたＣＰＵ２５０３、メモリ２５０１などのハードウエアとＯＳ及び必要なアプリケーション・プログラムとが有機的に協働することにより、上で述べたような各種機能を実現する。

【００８３】

（付記１）

文書更新前の検索インデックスを用いて検索を行うインデックス検索ユニットに対して、検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得する第１取得ステップと、

文書更新後の文書内容を保持する文字列検索ユニットに対して、前記検索要求に係る文字列検索を前記文書更新後の文書内容に対して実施させ、該当文書の文書ＩＤの第２のリストを取得する第２取得ステップと、

前記第１のリストと前記第２のリストと文書更新に係る文書の文書ＩＤの第３のリストとを用いて、前記検索要求に対する検索結果を生成する生成ステップと、

を含み、コンピュータにより実行される検索処理方法。

【００８４】

（付記２）

前記生成ステップが、

前記第１のリストから前記第３のリストに含まれる文書ＩＤを除去し、第４のリストを生成するステップと、

前記第４のリストと前記第２のリストとの和集合を生成するステップと、

を含む付記１記載の検索処理方法。

【００８５】

（付記３）

前記第１取得ステップが、前記検索要求に係る検索語に対してインデックス検索のための正規化処理を実施するステップを含み、

前記第２取得ステップが、前記検索要求に係る検索語に対して文字列検索のための正規化処理を実施するステップを含む

付記１記載の検索処理方法。

【００８６】

（付記４）

検索要求に応答してインデックス検索を実施するために用いられる第１の検索インデックスと、

検索時に用いられ且つ更新文書の文書ＩＤを格納する第１の差分文書ＩＤ管理部と、

検索時に文字列検索を実施するために前記更新文書のデータを格納する第１のパターンマッチング対象文書格納部と、

を有し、

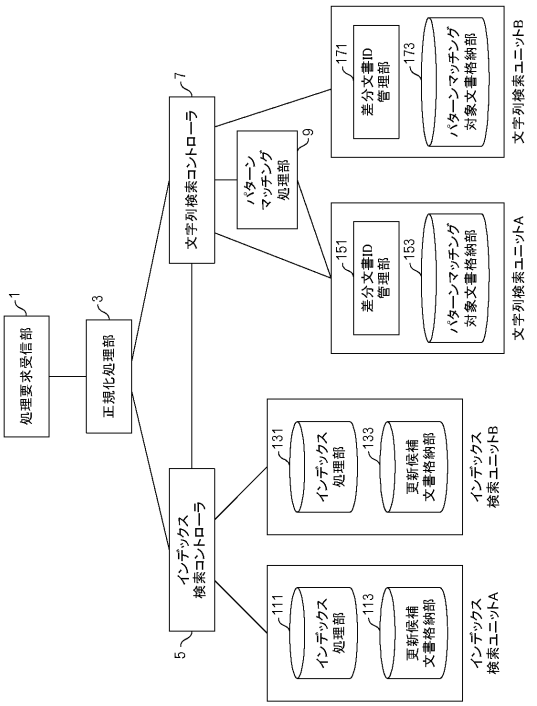
前記第１の検索インデックスを用いて前記検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得し、

前記検索要求に係る文字列検索を前記第１のパターンマッチング対象文書格納部に対して実施し、該当文書の文書ＩＤの第２のリストを取得し、
前記第１のリストと前記第２のリストと第１の差分文書ＩＤ管理部に格納されている文書ＩＤとを用いて、前記検索要求に対する検索結果を生成する
検索システム。
【００８７】
（付記５）
前記検索結果を生成する処理において、
前記第１のリストから前記第１の差分文書ＩＤ管理部に格納されている文書ＩＤを除去し、第３のリストを生成し、
前記第２のリストと前記第３のリストとの和集合を生成する、
付記４記載の検索システム。
【００８８】
（付記６）
検索要求に応答してインデックス検索を実施するために用いられる第１の検索インデックスと、
更新文書を反映するために用いられる第２の検索インデックスと、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書のデータを格納する第１の更新候補文書格納部と、
前記第２の検索インデックスに対する反映処理に用いられる前記更新文書のデータを格納する第２の更新候補文書格納部と、
検索時に用いられ且つ前記更新文書の文書ＩＤを格納する第１の差分文書ＩＤ管理部と、
検索時に文字列検索を実施するために前記更新文書のデータを格納する第１のパターンマッチング対象文書格納部と、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書の文書ＩＤを格納する第２の差分文書ＩＤ管理部と、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書のデータを格納する第２のパターンマッチング対象文書格納部と、
を有し、
前記第２の検索インデックスに対する前記更新文書の反映処理が完了した場合、
前記反映処理後の前記第２の検索インデックスを、前記第１の検索インデックスとして切り替えて用い、
前記第１の更新候補文書格納部に格納されている前記更新文書のデータを前記第２の更新候補文書格納部に上書き移動させ、
前記第２の差分文書ＩＤ管理部に格納されている前記更新文書の文書ＩＤを前記第１の差分文書ＩＤ管理部に上書き移動させ、
前記第２のパターンマッチング対象文書格納部に格納されている前記更新文書のデータを前記第１のパターンマッチング対象文書格納部に上書き移動させる
検索システム。
【００８９】
（付記７）
前記第２の検索インデックスに対する前記更新文書の反映処理が開始される前、
前記第１及び第２の更新候補文書格納部に新たな更新要求に係る更新文書のデータを格納し、
前記第１の差分文書ＩＤ管理部に前記新たな更新要求に係る更新文書の文書ＩＤを格納し、
前記第１のパターンマッチング対象文書格納部に前記新たな更新要求に係る更新文書のデータを格納する
付記６記載の検索システム。

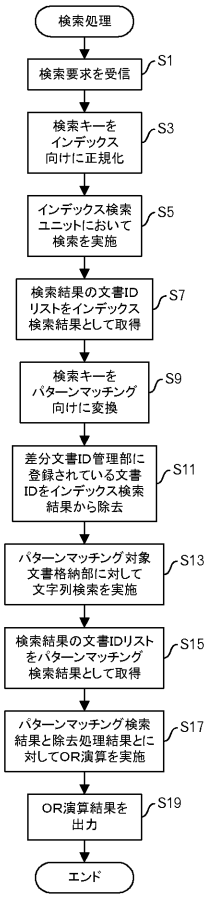
【００９０】
（付記８）
前記第２の検索インデックスに対する前記更新文書の反映処理が開始された後、
前記第１の更新候補文書格納部をクリアし、
前記第１の更新候補文書格納部に新たな更新要求に係る更新文書のデータを格納し、
前記第１及び第２の差分文書ＩＤ管理部に前記新たな更新要求に係る更新文書の文書ＩＤを格納し、
前記第１及び第２のパターンマッチング対象文書格納部に前記新たな更新要求に係る更新文書のデータを格納する
付記７記載の検索システム。
【００９１】
（付記９）
前記第１の検索インデックスを用いて前記検索要求に係る検索を実施させ、該当文書の文書ＩＤの第１のリストを取得し、
前記検索要求に係る文字列検索を前記第１のパターンマッチング対象文書格納部に対して実施し、該当文書の文書ＩＤの第２のリストを取得し、
前記第１のリストと前記第２のリストと第１の差分文書ＩＤ管理部に格納されている文書ＩＤとを用いて、前記検索要求に対する検索結果を生成する
付記６乃至８のいずれか１つ記載の検索システム。
【００９２】
（付記１０）
前記第１及び第２の更新候補文書格納部に格納される前記更新文書のデータが、検索インデックス向けに正規化されており、
前記第１及び第２のパターンマッチング対象文書格納部に格納される前記更新文書のデータが、文字列検索向けに正規化されている
付記６記載の検索システム。
【００９３】
（付記１１）
付記１乃至３のいずれか１つ記載の検索処理方法をコンピュータに実行させるためのプログラム。
【００９４】
（付記１２）
検索要求に応答してインデックス検索を実施するために用いられる第１の検索インデックスと、
更新文書を反映するために用いられる第２の検索インデックスと、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書のデータを格納する第１の更新候補文書格納部と、
前記第２の検索インデックスに対する反映処理に用いられる前記更新文書のデータを格納する第２の更新候補文書格納部と、
検索時に用いられ且つ前記更新文書の文書ＩＤを格納する第１の差分文書ＩＤ管理部と、
検索時に文字列検索を実施するために前記更新文書のデータを格納する第１のパターンマッチング対象文書格納部と、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書の文書ＩＤを格納する第２の差分文書ＩＤ管理部と、
前記第２の検索インデックスに対する前記更新文書の反映処理を開始してからの更新文書のデータを格納する第２のパターンマッチング対象文書格納部と、
を有する検索システムに対して、
前記第２の検索インデックスに対する前記更新文書の反映処理が完了した場合、
前記反映処理後の前記第２の検索インデックスを、前記第１の検索インデックスとして

切り替えて用い、
前記第１の更新候補文書格納部に格納されている前記更新文書のデータを前記第２の更新候補文書格納部に上書き移動させ、
前記第２の差分文書ＩＤ管理部に格納されている前記更新文書の文書ＩＤを前記第１の差分文書ＩＤ管理部に上書き移動させ、
前記第２のパターンマッチング対象文書格納部に格納されている前記更新文書のデータを前記第１のパターンマッチング対象文書格納部に上書き移動させる
処理を実施させるためのプログラム。
【図面の簡単な説明】
【００９５】
【図１】本発明の実施の形態におけるシステム概要図である。
【図２】検索処理の処理フローを示す図である。
【図３】インデックス差分更新開始処理の処理フローを示す図である。
【図４】更新時処理の処理フローの第１の部分を示す図である。
【図５】インデックス検索ユニットＡの初期状態を示す図である。
【図６】インデックス検索ユニットＢの初期状態を示す図である。
【図７】インデックス検索ユニットＡの状態を示す図である。
【図８】インデックス検索ユニットＢの状態を示す図である。
【図９】更新時処理の処理フローの第２の部分を示す図である。
【図１０】文字列検索ユニットＡの初期状態を示す図である。
【図１１】文字列検索ユニットＢの初期状態を示す図である。
【図１２】文字列検索ユニットＡの状態を示す図である。
【図１３】文字列検索ユニットＢの状態を示す図である。
【図１４】インデックス検索ユニットＡの状態を示す図である。
【図１５】インデックス検索ユニットＢの状態を示す図である。
【図１６】インデックス検索ユニットＡの状態を示す図である。
【図１７】インデックス検索ユニットＢの状態を示す図である。
【図１８】文字列検索ユニットＡの状態を示す図である。
【図１９】文字列検索ユニットＢの状態を示す図である。
【図２０】インデックス検索ユニットＡの状態を示す図である。
【図２１】文字列検索ユニットＡの状態を示す図である。
【図２２】文字列検索ユニットＢの状態を示す図である。
【図２３】差分更新終了処理の処理フローを示す図である。
【図２４】インデックス検索ユニットＡ及びＢの状態を示す図である。
【図２５】文字列検索ユニットＡ及びＢの状態を示す図である。
【図２６】正規化処理の処理フローを示す図である。
【図２７】コンピュータの機能ブロック図である。
【符号の説明】
【００９６】
１　処理要求受信部
３　正規化処理部
５　インデックス検索コントローラ
７　文字列検索コントローラ
９　パターンマッチング処理部
１１１，１３１　インデックス処理部
１１３，１３３　更新候補文書格納部
１５１，１７１　差分文書ＩＤ管理部
１５３，１７３　パターンマッチング対象文書格納部

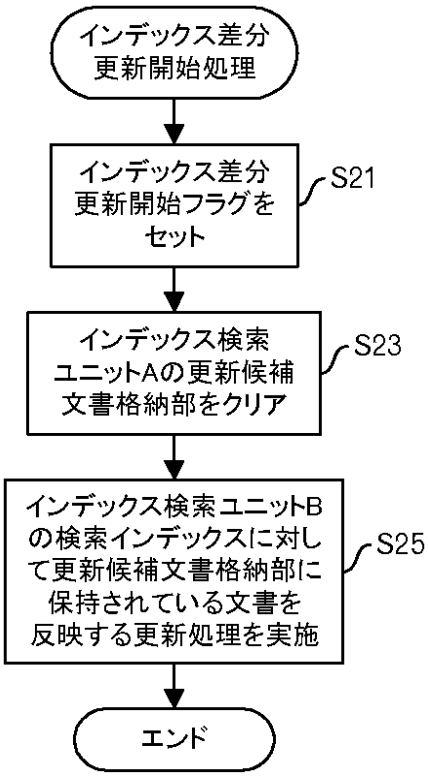
【図１】


処理要求受信部
1
正規化処理部
3
文字列検索コントローラ
7
パターンマッチング処理部
9
インデックス検索コントローラ
5
差分文書ID管理部
171
パターンマッチング対象文書格納部
173
文字列検索ユニットB
差分文書ID管理部
151
パターンマッチング対象文書格納部
153
文字列検索ユニットA
インデックス処理部
131
更新候補文書格納部
133
インデックス検索ユニットB
インデックス処理部
111
更新候補文書格納部
113
インデックス検索ユニットA

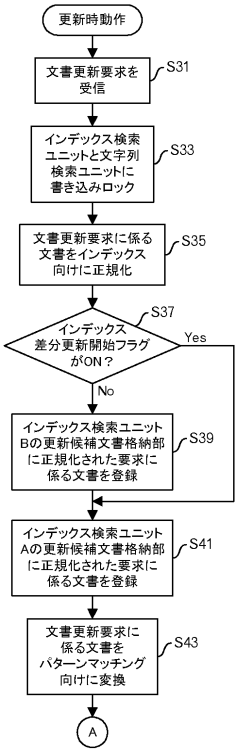
【図２】


検索処理
検索要求を受信
S1
検索キーをインデックス向けに正規化
S3
インデックス検索ユニットにおいて検索を実施
S5
検索結果の文書IDリストをインデックス検索結果として取得
S7
検索キーをパターンマッチング向けに変換
S9
差分文書ID管理部に登録されている文書IDをインデックス検索結果から除去
S11
パターンマッチング対象文書格納部に対して文字列検索を実施
S13
検索結果の文書IDリストをパターンマッチング検索結果として取得
S15
パターンマッチング検索結果と除去処理結果とに対してOR演算を実施
S17
OR演算結果を出力
S19
エンド

【図３】


インデックス差分更新開始処理
インデックス差分更新開始フラグをセット
S21
インデックス検索ユニットAの更新候補文書格納部をクリア
S23
インデックス検索ユニットBの検索インデックスに対して更新候補文書格納部に保持されている文書を反映する更新処理を実施
S25
エンド

【図４】


更新時動作
文書更新要求を受信
S31
インデックス検索ユニットと文字列検索ユニットに書き込みロック
S33
文書更新要求に係る文書をインデックス向けに正規化
S35
インデックス差分更新開始フラグがON？
S37
Yes
No
インデックス検索ユニットBの更新候補文書格納部に正規化された要求に係る文書を登録
S39
インデックス検索ユニットAの更新候補文書格納部に正規化された要求に係る文書を登録
S41
文書更新要求に係る文書をパターンマッチング向けに変換
S43
A

【図５】

インデックス処理部	空
更新候補文書格納部	空

インデックス検索ユニットA

【図６】

インデックス処理部	空
更新候補文書格納部	空

インデックス検索ユニットB

【図７】

インデックス処理部　空

更新候補文書格納部

文書ID	本文
00001	this is a book
00002	this is a pen
00003	this was a book
00004	that is a book
00005	that was a pen

インデックス検索ユニットA

【図８】

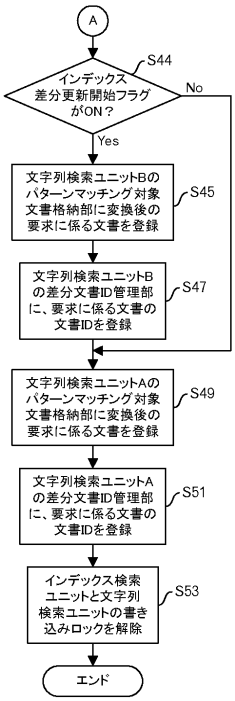
インデックス処理部　空

更新候補文書格納部

文書ID	本文
00001	this is a book
00002	this is a pen
00003	this was a book
00004	that is a book
00005	that was a pen

インデックス検索ユニットB

【図９】



【図１０】

差分文書ID管理部	空
パターンマッチング対象文書格納部	空

文字列検索ユニットA

【図１１】

差分文書ID管理部	空
パターンマッチング対象文書格納部	空

文字列検索ユニットB

【図１２】

差分文書ID管理部　00001, 00002, 00003, 00004, 00005

パターンマッチング対象文書格納部

文書ID	本文
00001	#this# #is# #a# #book#
00002	#this# #is# #a# #pen#
00003	#this# #was# #a# #book#
00004	#that# #is# #a# #book#
00005	#that# #was# #a# #pen#

文字列検索ユニットA

【図１３】

差分文書ID管理部	空
パターンマッチング対象文書格納部	空

文字列検索ユニットB

【図１４】

インデックス処理部	空
更新候補文書格納部	空

インデックス検索ユニットA

【図１５】

インデックス処理部　（更新開始）

更新候補文書格納部

文書ID	本文
00001	this is a book
00002	this is a pen
00003	this was a book
00004	that is a book
00005	that was a pen

インデックス検索ユニットB

【図１６】

インデックス処理部　空

更新候補文書格納部

文書ID	本文
00001	those are books

インデックス検索ユニットA

【図１７】

インデックス処理部　（更新中）

更新候補文書格納部

文書ID	本文
00001	this is a book
00002	this is a pen
00003	this was a book
00004	that is a book
00005	that was a pen

インデックス検索ユニットB

【図１８】

差分文書ID管理部　00001, 00002, 00003, 00004, 00005

パターンマッチング対象文書格納部

文書ID	本文
00001	#those# #are# #books#
00002	#this# #is# #a# #pen#
00003	#this# #was# #a# #book#
00004	#that# #is# #a# #book#
00005	#that# #was# #a# #pen#

文字列検索ユニットA

【図１９】

差分文書ID管理部　00001

パターンマッチング対象文書格納部

文書ID	本文
00001	#those# #are# #books#

文字列検索ユニットB

【図２０】

インデックス処理部　空

更新候補文書格納部

文書ID	本文
00001	those are books
00007	that is a pencil

インデックス検索ユニットA

【図２１】

差分文書ID管理部　00001, 00002, 00003, 00004, 00005, 00007

パターンマッチング対象文書格納部

文書ID	本文
00001	#those# #are# #books#
00002	#this# #is# #a# #pen#
00003	#this# #was# #a# #book#
00004	#that# #is# #a# #book#
00005	#that# #was# #a# #pen#
00007	#that# #is# #a# #pencil#

文字列検索ユニットA

【図２２】

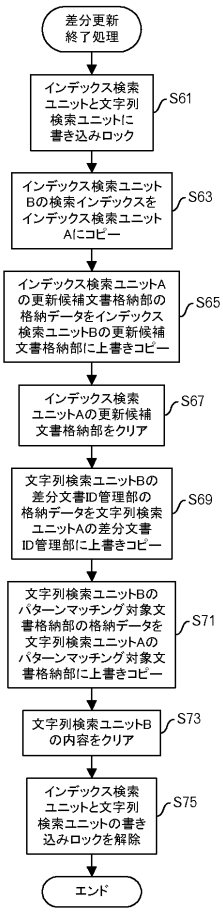
差分文書ID管理部　00001, 00007

パターンマッチング対象文書格納部

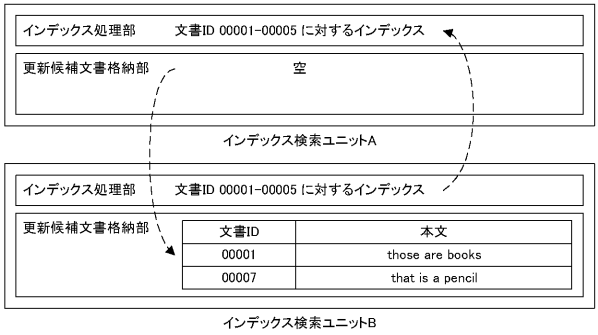
文書ID	本文
00001	#those# #are# #books#
00007	#that# #is# #a# #pencil#

文字列検索ユニットB

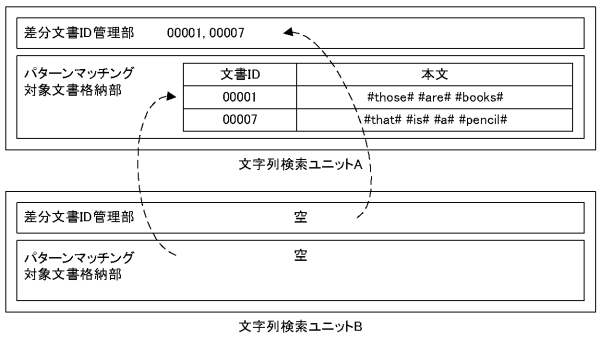
【図２３】



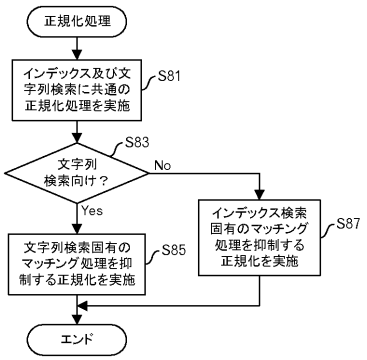
【図２４】



【図２５】



【図２６】



【図２７】

