

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4806773号
(P4806773)

(45) 発行日 平成23年11月2日(2011.11.2)

(24) 登録日 平成23年8月26日(2011.8.26)

(51) Int.Cl. F I
G06F 17/30 (2006.01)
 G06F 17/30 220A
 G06F 17/30 170A
 G06F 17/30 340Z
 G06F 17/30 350C

請求項の数 5 (全 41 頁)

<p>(21) 出願番号 特願2006-256759 (P2006-256759) (22) 出願日 平成18年9月22日 (2006. 9. 22) (65) 公開番号 特開2008-77459 (P2008-77459A) (43) 公開日 平成20年4月3日 (2008. 4. 3) 審査請求日 平成21年7月28日 (2009. 7. 28)</p> <p>特許法第30条第1項適用 2006年3月24日 国立大学法人 横浜国立大学大学院 環境情報学院発行の「環境情報からのメッセージVol. 6 博士・修士論文研究概要」に発表</p>	<p>(73) 特許権者 504182255 国立大学法人横浜国立大学 神奈川県横浜市保土ヶ谷区常盤台79番1号 (74) 代理人 100111800 弁理士 竹内 三明 (72) 発明者 森 辰則 神奈川県横浜市保土ヶ谷区常盤台79番1号 国立大学法人横浜国立大学内 (72) 発明者 谷岡 洵 神奈川県横浜市保土ヶ谷区常盤台79番1号 国立大学法人横浜国立大学内 審査官 宮下 浩次</p>
---	--

最終頁に続く

(54) 【発明の名称】 対話型複数文書要約装置

(57) 【特許請求の範囲】

【請求項1】

複数の文書を記憶する文書データベースから文書を検索する文書検索エンジンと接続する対話型複数文書要約装置であって、以下の要素を有することを特徴とする対話型複数文書要約装置

- (1) 検索条件文を入力する検索条件入力部
- (2) 入力した検索条件文から検索条件式を生成し、文書検索エンジンに検索条件式を転送し、文書検索エンジンから検索結果文書の識別情報を取得し、それらを要約対象文書の識別情報とする要約対象文書選択部
- (3) 要約対象文書の識別情報を記憶する要約対象文書記憶部
- (4) 要約対象文書の識別情報により特定される要約対象文書に含まれる文データに含まれる単語の重要度を算出し、当該単語の重要度に基づいて文の重要度を算出し、文の重要度に基づいて重要文を特定し、重要文からなる要約文書を生成する要約文書生成部
- (5) 要約文書を表示し、操作者に対して要約文書に含まれる重要文から注目文を選択するように促す要約文書画面を生成し、要約文書画面で選択された注目文の出所である注目文書の識別情報を特定する要約文書提示部
- (6) 要約対象文書記憶部に記憶している要約対象文書の識別情報群を注目文書の識別情報に更新する要約対象文書絞込部。

【請求項2】

要約文書提示部は、要約文書画面が操作者に対して要約文書から注目文字列を選択する

ように促すように生成し、要約文書画面で選択された注目文字列に含まれる注目文字列内単語を特定し、更に要約文書に含まれるが注目文字列に含まれない注目文字列外単語を特定し、

対話型複数文書要約装置は、注目文字列内単語に1より大きい文字列注目係数を設定し、注目文字列外単語に1より小さい文字列注目係数を設定する文字列注目係数算出部を有し、

要約文書生成部は、単語の重要度に文字列注目係数を乗じ、その積を単語の重要度として用いることを特徴とする請求項1記載の対話型複数文書要約装置。

【請求項3】

要約文書提示部は、更に重要文の出所のうち注目文書以外の非注目文書の識別情報を特定し、

対話型複数文書要約装置は、更に、注目文書に含まれる注目文書内単語に1より大きい文書注目係数を設定し、非注目文書に含まれる単語であって注目文書内単語以外の注目文書外単語に1より小さい文書注目係数を設定する文書注目係数算出部を有し、

要約対象要約文書生成部は、単語の重要度に文書注目係数を乗じ、その積を単語の重要度として用いることを特徴とする請求項1記載の対話型複数文書要約装置。

【請求項4】

対話型複数文書要約装置は、更に、

検索条件文に含まれる単語の頻度に基づいて初期の検索質問ベクトルを算出し、検索質問ベクトル記憶部に記憶させる初期検索質問ベクトル算出部と、

検索質問ベクトルを記憶する検索質問ベクトル記憶部と、

文書検索エンジンに検索質問ベクトルを転送し、文書検索エンジンから検索結果文書の識別情報を取得し、それらを推薦要約対象文書の識別情報とする推薦要約対象文書抽出部と、

推薦要約対象文書の識別情報により特定される推薦要約対象文書に含まれる文データに含まれる単語の重要度を算出し、当該単語の重要度に基づいて文の重要度を算出し、文の重要度に基づいて重要文を特定し、重要文からなる推薦要約文書を生成する推薦要約文書生成部と、

注目文書の識別情報により特定される文書ベクトルに基づいて加算修正ベクトルを算出し、非注目文書の識別情報により特定される文書ベクトルに基づいて減算修正ベクトルを算出し、検索質問ベクトル記憶部に記憶している検索質問ベクトルに加算修正ベクトルを加算し、減算修正ベクトルを減算し、検索質問ベクトルを更新する検索質問ベクトル更新部を有し、

要約文書提示部は、推薦要約文書を表示し、操作者に対して推薦要約文書に含まれる重要文から注目文を選択するように促す推薦要約文書画面を生成し、推薦要約文書画面で選択された注目文の出所である注目文書の識別情報を特定し、前記要約文書及び当該推薦要約文書について、重要文の出所のうち注目文書以外の非注目文書の識別情報を特定することを特徴とする請求項1記載の対話型複数文書要約装置。

【請求項5】

要約対象文書の識別情報を記憶する要約対象文書記憶部を有し、複数の文書を記憶する文書データベースから文書を検索する文書検索エンジンと接続する対話型複数文書要約装置となるコンピュータに、以下の手順を実行させるためのプログラム

(1) 検索条件文を入力する検索条件入力手順

(2) 入力した検索条件文から検索条件式を生成し、文書検索エンジンに検索条件式を転送し、文書検索エンジンから検索結果文書の識別情報を取得し、それらを要約対象文書の識別情報とする要約対象文書選択手順

(3) 要約対象文書の識別情報により特定される要約対象文書に含まれる文データに含まれる単語の重要度を算出し、当該単語の重要度に基づいて文の重要度を算出し、文の重要度に基づいて重要文を特定し、重要文からなる要約文書を生成する要約文書生成手順

(4) 要約文書を表示し、操作者に対して要約文書に含まれる重要文から注目文を選択す

10

20

30

40

50

るように促す要約文書画面を生成し、要約文書画面で選択された注目文の出所である注目文書の識別情報を特定する要約文書提示手順

(5) 要約対象文書記憶部に記憶している要約対象文書の識別情報群を注目文書の識別情報に更新する要約対象文書絞込手順。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索条件により検索した複数の文書を要約する対話型複数文書要約装置に係り、要約文書から操作者が選択する注目部位に基づいて、注目部位に関連する情報を含む要約文書を再生成する技術に関する。

10

【背景技術】

【0002】

通常、要約生成装置では、複数の文書から重要文を抽出し、それらを整列させて要約文書を生成する。しかし、一回の要約で操作者が欲する情報を提供できるとは限らない。

【特許文献1】特開2006-59082号公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

本発明では、少ない操作と処理により、目的の情報に容易に到達できるようにすることを課題とする。

20

【課題を解決するための手段】

【0004】

本発明に係る対話型複数文書要約装置は、

複数の文書を記憶する文書データベースから文書を検索する文書検索エンジンと接続する対話型複数文書要約装置であって、以下の要素を有することを特徴とする

(1) 検索条件文を入力する検索条件入力部

(2) 入力した検索条件文から検索条件式を生成し、文書検索エンジンに検索条件式を転送し、文書検索エンジンから検索結果文書の識別情報を取得し、それらを要約対象文書の識別情報とする要約対象文書選択部

30

(3) 要約対象文書の識別情報を記憶する要約対象文書記憶部

(4) 要約対象文書の識別情報により特定される要約対象文書に含まれる文データに含まれる単語の重要度を算出し、当該単語の重要度に基づいて文の重要度を算出し、文の重要度に基づいて重要文を特定し、重要文からなる要約文書を生成する要約文書生成部

(5) 要約文書を表示し、操作者に対して要約文書に含まれる重要文から注目文を選択するように促す要約文書画面を生成し、要約文書画面で選択された注目文の出所である注目文書の識別情報を特定する要約文書提示部

(6) 要約対象文書記憶部に記憶している要約対象文書の識別情報群を注目文書の識別情報に更新する要約対象文書絞込部。

【0005】

40

また、要約文書提示部は、要約文書画面が操作者に対して要約文書から注目文字列を選択するように促すように生成し、要約文書画面で選択された注目文字列に含まれる注目文字列内単語を特定し、更に要約文書に含まれるが注目文字列に含まれない注目文字列外単語を特定し、

対話型複数文書要約装置は、注目文字列内単語に1より大きい文字列注目係数を設定し、注目文字列外単語に1より小さい文字列注目係数を設定する文字列注目係数算出部を有し、

要約文書生成部は、単語の重要度に文字列注目係数を乗じ、その積を単語の重要度として用いることを特徴とする。

【0006】

50

また、要約文書提示部は、更に重要文の出所のうち注目文書以外の非注目文書の識別情報を特定し、

対話型複数文書要約装置は、更に、注目文書に含まれる注目文書内単語に1より大きい文書注目係数を設定し、非注目文書に含まれる単語であって注目文書内単語以外の注目文書外単語に1より小さい文書注目係数を設定する文書注目係数算出部を有し、

要約対象要約文書生成部は、単語の重要度に文書注目係数を乗じ、その積を単語の重要度として用いることを特徴とする。

【0007】

また、対話型複数文書要約装置は、更に、

検索条件文に含まれる単語の頻度に基づいて初期の検索質問ベクトルを算出し、検索質問ベクトル記憶部に記憶させる初期検索質問ベクトル算出部と、

検索質問ベクトルを記憶する検索質問ベクトル記憶部と、

文書検索エンジンに検索質問ベクトルを転送し、文書検索エンジンから検索結果文書の識別情報を取得し、それらを推薦要約対象文書の識別情報とする推薦要約対象文書抽出部と、

推薦要約対象文書の識別情報により特定される推薦要約対象文書に含まれる文データに含まれる単語の重要度を算出し、当該単語の重要度に基づいて文の重要度を算出し、文の重要度に基づいて重要文を特定し、重要文からなる推薦要約文書を生成する推薦要約文書生成部と、

注目文書の識別情報により特定される文書ベクトルに基づいて加算修正ベクトルを算出し、非注目文書の識別情報により特定される文書ベクトルに基づいて減算修正ベクトルを算出し、検索質問ベクトル記憶部に記憶している検索質問ベクトルに加算修正ベクトルを加算し、減算修正ベクトルを減算し、検索質問ベクトルを更新する検索質問ベクトル更新部を有し、

要約文書提示部は、推薦要約文書を表示し、操作者に対して推薦要約文書に含まれる重要文から注目文を選択するように促す推薦要約文書画面を生成し、推薦要約文書画面で選択された注目文の出所である注目文書の識別情報を特定し、前記要約文書及び当該推薦要約文書について、重要文の出所のうち注目文書以外の非注目文書の識別情報を特定することを特徴とする。

【0008】

本発明に係るプログラムは、

要約対象文書の識別情報を記憶する要約対象文書記憶部を有し、複数の文書を記憶する文書データベースから文書を検索する文書検索エンジンと接続する対話型複数文書要約装置となるコンピュータに、以下の手順を実行させることを特徴とする

(1) 検索条件文を入力する検索条件入力手順

(2) 入力した検索条件文から検索条件式を生成し、文書検索エンジンに検索条件式を転送し、文書検索エンジンから検索結果文書の識別情報を取得し、それらを要約対象文書の識別情報とする要約対象文書選択手順

(3) 要約対象文書の識別情報により特定される要約対象文書に含まれる文データに含まれる単語の重要度を算出し、当該単語の重要度に基づいて文の重要度を算出し、文の重要度に基づいて重要文を特定し、重要文からなる要約文書を生成する要約文書生成手順

(4) 要約文書を表示し、操作者に対して要約文書に含まれる重要文から注目文を選択するように促す要約文書画面を生成し、要約文書画面で選択された注目文の出所である注目文書の識別情報を特定する要約文書提示手順

(5) 要約対象文書記憶部に記憶している要約対象文書の識別情報群を注目文書の識別情報に更新する要約対象文書絞込手順。

【発明の効果】

【0009】

本発明によれば、操作者が選択した注目文の出所である注目文書を要約対象とするので、操作者が関心を持った情報の源から有効な要約文書を生成することができる。

【 0 0 1 0 】

また、操作者が選択した文字列に含まれる単語の重要度を増して重要文を選択するので、操作者が関心を持った情報と直接的に関連の深い文を抽出することができる。

【 0 0 1 1 】

また、操作者が選択した注目文の出所である注目文書に含まれる単語の重要度を増して重要文を選択するので、操作者が関心を持った情報の源と関連の深い文を抽出することができる。

【 0 0 1 2 】

ベクトルによる文書検索で生成した推薦要約を併せて表示し、推薦要約からも注目文書を選択させるので、要約対象の文書の収束を防止することができる。

10

【 発明を実施するための最良の形態 】

【 0 0 1 3 】

実施の形態 1 .

本実施の形態では、表示された要約文書中から操作者が選択した文に基づいて、当該文の出所である文書に要約の対象を絞り込む実施例を説明する。

【 0 0 1 4 】

図 1 は、実施の形態 1 に係る全体処理フローを示す図である。検索条件入力部 2 0 1 による検索条件入力処理 (S 1 0 1) で、単語を含む検索条件文の入力を受付ける。要約対象文書選択部 2 0 2 による要約対象文書選択処理 (S 1 0 2) では、検索条件文に基づく検索条件式に従って文書データベース 2 0 4 内の文書を文書検索エンジン 2 0 3 に検索させ、検索された文書を要約対象文書とする。本処理については、図 3 を用いて後述する。要約文書生成部 5 0 1 による要約文書生成処理 (S 1 0 3) では、要約対象文書に含まれる文の重要度を判定し、文重要度の大きい文群からなる要約文書を生成する。本処理については、図 7 を用いて後述する。要約文書提示部 5 0 4 による要約文書提示処理 (S 1 0 4) では、生成した要約文書を表示するとともに、操作者に注目する部位 (本形態では、文) を指示させる。本処理については、図 2 5 を用いて後述する。要約文書提示処理 (S 1 0 4) の終了ステータス (S 1 0 5) として絞込要約ステータスを受け付けた場合に、要約対象文書絞込部 5 0 6 による要約対象絞込処理 (S 1 0 6) で前述の注目部位に従って要約文書を絞り込む。本処理については、図 2 9 を用いて後述する。終了ステータス (S 1 0 5) として完了ステータスを受け付けた場合には、全体処理を終了する。

20

30

【 0 0 1 5 】

前述の検索条件入力処理 (S 1 0 1) 及び要約対象文書選択処理 (S 1 0 2) について詳述する。図 2 は、要約対象文書選択に係るモジュール構成を示す図である。対話型複数文書要約装置は、検索条件入力部 2 0 1、要約対象文書選択部 2 0 2、文書検索エンジン 2 0 3、文書データベース 2 0 4、及び要約対象文書記憶部 2 0 5 を有する。

【 0 0 1 6 】

図 3 は、要約対象文書選択処理フローを示す図である。要約対象文書選択部 2 0 2 は、検索条件文に従って、検索条件式を生成する (S 3 0 1)。そして、文書検索エンジン 2 0 3 に検索条件式を転送し文書を検索させ (S 3 0 2)、文書検索エンジン 2 0 3 から検索結果の文書 ID 群を取得する (S 3 0 3)。取得した文書 ID 群を要約対象文書記憶部 2 0 5 に記憶させる (S 3 0 4)。

40

【 0 0 1 7 】

図 4 は、要約対象文書記憶部の構成例を示す図である。検索結果の文書 ID 群を要約対象の文書 ID 群として記憶している。

【 0 0 1 8 】

図 5 は、要約文書生成、要約文書提示、及び要約対象文書絞込に係るモジュール構成を示す図である。本装置は、要約文書生成部 5 0 1、要約文書記憶部 5 0 2、要約文書構成テーブル 5 0 3、要約文書提示部 5 0 4、注目文書テーブル 5 0 5、及び要約対象文書絞込部 5 0 6 を有している。

【 0 0 1 9 】

50

次に、要約文書生成部 501 による要約文書生成処理 (S103) について詳述する。図 6 は、要約文書生成部の内部モジュール構成を示す図である。文書データベース 204 は、文書管理テーブル 602、単語テーブル 603、文内単語出現頻度テーブル 604、文書内単語出現頻度テーブル 605、文書頻度逆数テーブル 606、文提供部 607、及び文書データ記憶部 608 を有している。要約文書生成部 501 は、文重要度計算部 601、文重要度テーブル 609、重要文抽出部 610、文抽出テーブル 611、及び重要文整列部 612 を有している。

【0020】

図 7 は、要約文書生成処理フローを示す図である。まず、文重要度テーブル 609、文抽出テーブル 611、要約文書記憶部 502、及び要約文書構成テーブル 503 をクリアする (S701)。文重要度計算部 601 による文重要度計算処理 (S702) では、要約対象文書に含まれる文の重要度を判定する。本処理については、図 8 を用いて後述する。重要文抽出部 610 による重要文抽出処理 (S703) では、文重要度に基づいて、要約文書に用いる文を抽出する。本処理については、図 18 を用いて後述する。重要文整列部 612 による重要文整列処理 (S704) では、抽出した文を整列させて要約文書を生成する。本処理については、図 21 を用いて後述する。

10

【0021】

文重要度計算部 601 による文重要度計算処理 (S702) について詳述する。図 8 及び図 9 は、文重要度計算処理フローを示す図である。要約対象文書記憶部 205 に記憶している要約対象文書毎に (S801)、当該要約対象文書に含まれる文毎の処理を繰り返す (S802)。文の数は、文書管理テーブル 602 から取得する。

20

【0022】

図 10 に、文書管理テーブルの構成例を示す。文書管理テーブル 602 では、文書毎にレコードを設け、文書 ID と、文書作成日付と、文書データファイル名と、文書を構成する文数と、文書内単語総数の項目を対応付けて記憶するように構成されている。

【0023】

文毎の処理として、当該文を解析し、当該文に含まれる単語群を特定し (S803)、特定した各単語に対応する単語 ID を特定する (S804)。これらの対応付けは、単語テーブル 603 より取得する。

【0024】

図 11 は、単語テーブルの構成例を示す図である。単語毎にレコードを設け、単語 ID と、単語文字列との項目を対応付けて記憶するように構成されている。

30

【0025】

次に、当該文に含まれる総単語数 (文中総単語数) を算出する (S805)。この文中総単語数算出処理については、図 12 を用いて後述する。

【0026】

そして、当該文に含まれる単語毎に以下の処理を繰り返す (S806)。単語毎の処理として、当該単語の重要度を算出する (S807)。この単語重要度算出処理については、図 14 を用いて後述する。そして、算出した単語重要度を累計パラメータ (初期値 = 0) に加算する (S808)。これらの処理を、当該文に含まれるすべての単語について処理した後 (S809)、累計パラメータの値 (単語重要度の総計) を文中総単語数で除算して、商を得る (S810)。そして、当該文書の文書 ID 及び当該文の文書内文番号と対応付けて、算出した商を文重要度テーブルに一時的に記憶させる (S811)。このようにして、当該要約対象文書に含まれるすべての文について処理すると (S812)、最後に当該要約対象文書に含まれる文群の前記商を母集団として、各商の偏差値 (T-score, 0 ~ 1 の値) を算出し、算出した偏差値を文重要度として文重要度テーブルを書き変える (S813)。そして、すべての要約対象文書について処理した時点で終了する (S814)。

40

【0027】

ここで、前述の文中総単語数算出処理 (S805) について詳述する。図 12 は、文中

50

総単語数算出処理フローを示す図である。当該文に含まれる単語毎に以下の処理を繰り返す（S1201）。文内単語出現頻度テーブル604から当該文書の当該文中における当該単語の出現頻度（文内単語出現頻度）を取得し（S1202）、文内単語出現頻度を累計パラメータ（初期値=0）に加算する（S1203）。そして、当該文に含まれるすべての単語について処理した時点で（S1204）、累計パラメータの値（文内単語出現頻度の総計）を文中総単語数とする（S1205）。

【0028】

文内単語出現頻度テーブル604は、図13のようになっている。文書IDに対応付けられ文書毎に設けられたテーブルが文書数分設けられており、各テーブルは、当該文書の文毎に各単語の出現頻度を記憶するように構成されている。

10

【0029】

更に、単語重要度算出処理（S807）について詳述する。図14は、単語重要度算出処理フローを示す図である。文内単語出現頻度テーブル604から当該文書の当該文中における当該単語の出現頻度（文内単語出現頻度）を取得し（S1401）、更に文書内単語出現頻度テーブル605から当該文書における当該単語の出現頻度（文書内単語出現頻度）を取得する（S1402）。

【0030】

文書内単語出現頻度テーブル605は、図15のようになっている。当該文書毎に各単語の出現頻度を記憶するように構成されている。

【0031】

また、文書頻度逆数テーブル606から当該単語の文書頻度の逆数（文書頻度逆数）を取得する（S1403）。

20

【0032】

図16は、文書頻度逆数テーブルの構成例を示す図である。単語毎に、文書頻度の逆数（IDF値）を記憶するように構成されている。

【0033】

そして、文内単語出現頻度と文書内単語出現頻度と文書頻度逆数を乗算し（S1404）、算出した積を単語重要度として文重要度テーブル609に記憶させる（S1405）。

【0034】

図17は、文重要度テーブルの構成例を示す図である。要約対象文書に含まれる文毎にレコードを設け、当該要約対象文書の文書IDと、その文書内の文番号と、その文の重要度との項目を対応付けて記憶するように構成されている。

30

【0035】

次に、重要文抽出部610による重要文抽出処理（S703）について詳述する。図18と図19は、重要文抽出処理フローを示す図である。この処理において、抽出文テーブルを用いる。

【0036】

図20は、文抽出テーブルの構成例を示す図である。要約対象文書に含まれる文毎にレコードを設け、当該要約対象文書の文書IDと、その文書内の文番号と、抽出フラグとの項目を対応付けて記憶するように構成されている。抽出された文を、ONとして識別する。

40

【0037】

まず、既選択文集合Aを空集合に初期化する（S1801）。具体的には、抽出文テーブルのすべての抽出フラグをOFFにする。次に、変数である既選択文長Lを0に初期化する（S1802）。

【0038】

そして、要約対象の文書に含まれる文の集合SSと既選択文集合Aの差集合に含まれる文（Sp = SS \ A）毎に以下の処理を行う（S1803）。具体的には、抽出文テーブルの抽出フラグがOFFの文について処理する。

50

【0039】

既選択文集合 A に含まれる文 (S q A) 毎に以下の処理を繰り返す (S 1 8 0 4) 。
 具体的には、抽出文テーブルの抽出フラグが ON の文について処理する。差集合に含まれる文 (S p) と既選択文集合に含まれる文 (S q) の類似度 (S i m_s (S p , S q)) を算出する (S 1 8 0 5) 。既選択文集合に含まれる文 (S q) のすべてについて処理した時点で (S 1 8 0 6) 、次に移行する。

【0040】

既選択文集合に含まれる各文との組み合わせによる類似度のうち、最大の類似度 (m a x S i m_s (S p , S q)) を選択し (S 1 8 0 7) 、最大の類似度に、(1 - 冗長度制御変数) を乗じて積を求め、積を第二項の値 ((1 -) m a x S i m_s (S p , S q)) とする (S 1 8 0 8) 。

10

【0041】

文重要度テーブル 6 0 9 から、差集合に含まれる文 (S p) の文重要度 (I m p_s (S p)) を読み込み (S 1 8 0 9) 、文重要度に冗長度制御変数 を乗じて積を求め、積を第一項の値 (I m p_s (S p)) とする (S 1 8 1 0) 。

【0042】

そして、第一項の値から第二項の値を引いて差を求め、差を抽出評価値とする (S 1 8 1 1) 。差集合に含まれる文 (S p) のすべてについて処理すると (S 1 8 1 2) 、差集合に含まれる文 (S p) のうち、前記抽出評価値が最大となる文 (S p) を特定する (S 1 8 1 3) 。

20

【0043】

前記抽出評価値が最大の文の長さ (| S p |) を既選択文長 L に加え (S 1 8 1 4) 、既選択文長 L が要約文書制限長を越えた場合には (S 1 8 1 5) 、終了する。越えていない場合には、前記抽出評価値が最大の文 (S p) を既選択文集合 A に加える (S 1 8 1 6) 。具体的には、抽出文テーブルの当該文の文 ID に対応する抽出フラグを ON にする。そして、処理を繰り返す。

【0044】

次に、重要文整列部 6 1 2 による重要文整列処理 (S 7 0 4) について詳述する。図 2 1 は、重要文整列処理フローを示す図である。最初に、要約対象の文書について非階層型のクラスタリングを行なう (S 2 1 0 1) 。そして、そのクラスタ間の順序付けを行なう (S 2 1 0 2) 。例えば、文書作成日付に基づいて順序付けを行なう。更に、クラスタ内の文書間の順序付けを行なう (S 2 1 0 3) 。

30

【0045】

そして、文書の順序付けに従って、文書毎に以下の処理を繰り返す (S 2 1 0 4) 。文抽出テーブルの抽出フラグが ON のレコードのうち、そのレコードに含まれる文書 ID が当該文書の ID と一致するレコードを検索し (S 2 1 0 5) 、検索したレコード群について、それらのレコードに含まれる文書内文番号の順に従ってレコード毎に以下の処理を繰り返す (S 2 1 0 6) 。文書データベース 2 0 4 の文提供部 6 0 7 から当該レコードに含まれる文書 ID と文書内文番号により特定される文データ (文字列) を取得し (S 2 1 0 7) 、取得した文データ (文字列) を、要約文書記憶部に順に記憶させる (S 2 1 0 8)

40

図 2 2 は、要約文書記憶部の構成例を示す図である。要約内文番号の順に従って重要文を記憶している。

【0046】

また、当該文書 ID と当該文書内文番号を対応付けて、要約文構成テーブル 5 0 3 に順に記憶させる (S 2 1 0 9) 。

【0047】

図 2 3 は、要約文書構成テーブルの構成例を示す図である。要約文書内の文毎にレコードを設け、要約内文番号の順に従って文書 ID と、その文書内の文番号の項目を対応付けて記憶するように構成されている。

【0048】

50

そして、すべてのレコードについて処理すると（S 2 1 1 0）、次の文書の処理に移り、最終的に順序付けしたすべての文書について処理した時点で終了する（S 2 1 1 1）。

【 0 0 4 9 】

次に、前述の要約文書提示部 5 0 4 による要約文書提示処理（S 1 0 4）について詳述する。

図 2 4 は、要約文書提示部の内部モジュール構成を示す図である。要約文書提示部 5 0 4 は、要約文書画面生成部 2 4 0 1、要約文書画面記憶部 2 4 0 2、要約文書画面出力部 2 4 0 3、注目部位取得部 2 4 0 4、注目文記憶部 2 4 0 5、及び注目文書判定部 2 4 0 6 を有している。

【 0 0 5 0 】

図 2 5 は、要約文書提示処理フローを示す図である。まず、注目文記憶部 2 4 0 5、及び注目文書テーブル 5 0 5 をクリアする（S 2 5 0 1）。そして、要約文書画面生成部 2 4 0 1 による要約文書画面生成処理（S 2 5 0 2）で、要約文書を表示し、要約文書中の注目部位（この例では、注目する文書）の選択などの指示を促す要約文書画面を生成する。本処理については、図 2 6 を用いて後述する。次に、要約文書画面出力部 2 4 0 3 による要約文書画面出力処理で（S 2 5 0 3）、生成した要約文書画面を出力する。つまり、表示のために表示装置に送信する。そして、要約文書画面上での操作による指示を受け付ける（S 2 5 0 4）。注目文を選択する指示の場合には、注目部位取得部 2 4 0 4 による注目文取得処理（S 2 5 0 5）で、選択された注目文を特定する。詳しくは、図 2 7 を用いて後述する。受け付けた指示が絞込を要求する指示の場合には、注目文書判定部 2 4 0 6 による注目文書判定処理（S 2 5 0 6）で、注目文書を特定し、本処理の終了ステータスを絞込要約ステータスとして終了する（S 2 5 0 7）。注目文書判定処理（S 2 5 0 6）については、図 2 8 を用いて後述する。受け付けた指示が完了の指示の場合には、本処理の終了ステータスを完了ステータスとして終了する（S 2 5 0 8）。

【 0 0 5 1 】

要約文書画面生成部 2 4 0 1 による要約文書画面生成処理（S 2 5 0 2）について詳述する。図 2 6 は、要約文書画面生成処理フローを示す図である。要約文書記憶部 5 0 2 から順次重要文の文データを読み出し、以下の処理を繰り返す（S 2 6 0 1）。まず、文データの要約文書内文番号を特定する。具体的には、要約文書内文番号（初期値 = 0）をインクリメントする（S 2 6 0 2）。そして、要約文書画面に含まれる新たな文イベントを設け（S 2 6 0 3）、文アイテムの表示データとして、文データを設定する（S 2 6 0 4）。また、文アイテムの返信データとして、要約文書内文番号を設定する（S 2 6 0 5）。これにより、操作者が画面上で当該文データを選択した場合に、要約文書内文番号が返信されるようになる。そして、すべての文データについて処理した時点で終了する（S 2 6 0 6）。

【 0 0 5 2 】

次に、注目部位取得部 2 4 0 4 による注目文取得処理（S 2 5 0 5）について詳述する。図 2 7 は、注目文取得処理フローを示す図である。操作者の操作により選択された文アイテムの返信として、要約文書内文番号を取得し（S 2 7 0 1）、取得した要約文書内文番号を、注目文記憶部 2 4 0 5 に追加記憶させる（S 2 7 0 2）。

【 0 0 5 3 】

次に、注目文書判定部 2 4 0 6 による注目文書判定処理（S 2 5 0 6）について詳述する。図 2 8 は、注目文書判定処理フローを示す図である。注目文記憶部 2 4 0 5 に記憶している要約文書内注目文番号を順次読み出し、以下の処理を繰り返す（S 2 8 0 1）。要約文書構成テーブル 5 0 3 から要約文書内注目文番号に対応する文書 ID を取得し（S 2 8 0 2）、取得した文書 ID を注目文書テーブル 5 0 5 に記憶させる（S 2 8 0 3）。そして、すべての要約文書内注目文番号について処理した時点で終了する（S 2 8 0 4）。

【 0 0 5 4 】

続いて、要約対象文書絞込部 5 0 6 による要約対象絞込処理（S 1 0 6）について詳述する。図 2 9 は、要約対象文書絞込処理フローを示す図である。注目文書テーブル 5 0 5

10

20

30

40

50

から文書ID群を読み取る(S2901)。要約対象文書記憶部205を、読み取った文書ID群に書き換える(S2902)。このようにして、要約対象文書が絞り込まれる。

【0055】

実施の形態2.

本実施の形態では、表示された要約文書中から操作者が選択した文字列に基づいて、当該文字列に含まれる単語に対して重み付けを施して、要約文書を再生成する実施例について説明する。この例では、単語に対して重み付けとしてBIASW(文字列注目係数)を用いる。

【0056】

図30は、実施の形態2に係る全体処理フローを示す図である。本実施の形態では、要約文書生成部501による要約文書生成処理(S3003)で、文字列注目係数を用いた文重要度の算出を行う。本処理については、図39を用いて後述する。また、要約文書提示処理(S3004)では、注目部位として注目する文字列の指示の入力を受け付ける。つまり、選択した文字列を返信するように構成された要約文書表示画面を生成する。また、文字列注目係数算出部3103による文字列注目係数算出処理(S3007)では、選択された文字列に従って、文字列注目係数を設定する処理を行なう。本処理については、図37を用いて後述する。

10

【0057】

図31は、要約文書提示及び文字列注目係数算出に係るモジュール構成を示す図である。対話型複数文書要約装置は、注目文字列内単語テーブル3101、注目文字列外単語テーブル3102、文字列注目係数算出部3103、及び文字列注目係数テーブル3104を有している。

20

【0058】

要約文書提示部504による要約文書提示処理(S3003)について詳述する。図32は、要約文書提示部の内部モジュール構成を示す図である。要約文書提示部504は、注目文字列記憶部3201、注目文字列内単語判定部3202、及び注目文字列外単語判定部3203を有している。

【0059】

要約文書提示部504による要約文書提示処理(S3003)について詳述する。図33は、要約文書提示処理フローを示す図である。まず、注目文記憶部2405、注目文書テーブル505、注目文字列記憶部3201、注目文字列内単語テーブル3101、及び注目文字列外単語テーブル3102をクリアする(S3301)。要約文書画面生成処理(S3302)では、選択した文字列を返信するように文アイテムを設定した要約文書表示画面を生成する。尚、前述の例と同様に当該文字列を含む要約文書内文番号も返信する。要約文書画面出力処理(S3303)では、当該画面を出力する。指示受付処理(S3304)で、注目する文字列が指示された場合には、注目部位取得部2404による注目文字列取得処理(S3305)で、選択された要約文書内文番号に併せて選択された文字列を取得する。本処理については、図34を用いて後述する。受け付けた指示が、絞込要求指示の場合には、注目文書判定部2406による注目文書判定処理(S3306)で、注目文書を判定し、注目文字列内単語判定部3202による注目文字列内単語判定処理(S3307)で、注目文字列に含まれる単語を判定する。本処理については、図35を用いて後述する。また、注目文字列外単語判定部3203による注目文字列外単語判定処理(S3308)で、要約文書に含まれるそれ以外の単語を判定する。本処理については、図36を用いて後述する。

30

40

【0060】

注目部位取得部2404による注目文字列取得処理(S3305)について詳述する。図34は、注目文字列取得処理フローを示す図である。要約文書から選択された文字列と、当該文字列を含む文の要約文書内文番号を返値として取得し(S3401)、取得した文字列を、注目文字列として注目文字列記憶部3201に追加記憶させ(S3402)、取得した要約文書内文番号を、注目文記憶部2405に追加記憶させる(S3403)。

50

【 0 0 6 1 】

注目文字列内単語判定部 3 2 0 2 による注目文字列内単語判定処理 (S 3 3 0 7) について詳述する。図 3 5 は、注目文字列内単語判定処理フローを示す図である。注目文字列記憶部 3 2 0 1 に記憶している注目文字列を順次読み出し、以下の処理を繰り返す (S 3 5 0 1)。注目文字列を解析し、注目文字列に含まれる単語群を抽出し (S 3 5 0 2)、抽出した単語毎に以下の処理を繰り返す (S 3 5 0 3)。単語毎の処理として、単語テーブル 6 0 3 から当該単語に対応する単語 ID を検索し (S 3 5 0 4)、当該単語 ID を注目単語の ID として注目文字列内単語テーブルに追加記憶させる (S 3 5 0 5)。但し、重複する単語 ID は除く。すべての単語について処理すると (S 3 5 0 6)、次の注目文字列の処理に移行し、すべての注目文字列について処理した時点で終了する (S 3 5 0 7)。

10

【 0 0 6 2 】

次に、注目文字列外単語判定部 3 2 0 3 による注目文字列外単語判定処理 (S 3 3 0 8) について詳述する。図 3 6 は、注目文字列外単語判定処理フローを示す図である。要約文書記憶部要約文書記憶部 5 0 2 から重要文の文データを順次読み出し、以下の処理を繰り返す (S 3 6 0 1)。文データを解析し、文データに含まれる単語群を抽出し (S 3 6 0 2)、抽出した単語毎に以下の処理を繰り返す (S 3 6 0 3)。単語テーブル 6 0 3 から当該単語に対応する単語 ID を検索する (S 3 6 0 4)。特定した単語 ID は、注目文字列内単語テーブル 3 1 0 1 に含まれているか判定し (S 3 6 0 5)、含まれていない場合に、当該単語 ID を注目文字列外単語の ID として注目文字列外単語テーブル 3 1 0 2 に追加記憶させる (S 3 6 0 6)。但し、重複する単語 ID は除く。すべての単語について処理すると (S 3 6 0 7)、次の重要文の文データの処理に移行し、すべての文データについて処理した時点で終了する (S 3 6 0 8)。

20

【 0 0 6 3 】

次に、文字列注目係数算出部 3 1 0 3 による文字列注目係数算出処理 (S 3 0 0 7) について詳述する。図 3 7 は、文字列注目係数算出処理フローを示す図である。注目文字列内単語テーブル 3 1 0 2 から注目文字列内単語の ID 群を読み出し (S 3 7 0 1)、すべての注目文字列内単語 ID について、注目文字列内単語 ID に注目文字列内単語に係る文字列注目係数 (1 . 0 より大きい数値、例えば、定数 2 . 0) を対応付けて文字列注目係数テーブル 3 1 0 4 に記憶させる (S 3 7 0 2)。また、注目文字列外単語テーブル 3 1 0 2 から注目文字列外単語の ID 群を読み出し (S 3 7 0 3)、すべての注目文字列外単語 ID について、注目文字列外単語 ID に注目文字列外単語に係る文字列注目係数 (1 . 0 より小さい数値、例えば、定数 0 . 5) を対応付けて文字列注目係数テーブル 3 1 0 4 に記憶させる (S 3 7 0 4)。

30

【 0 0 6 4 】

続いて、要約文書生成処理 (S 3 0 0 3) 内の文重要度計算処理 (S 7 0 2) について詳述する。図 3 8 は、文重要度計算に係るモジュール構成を示す図である。図 3 9 は、単語重要度算出処理フローを示す図である。S 3 9 0 5 までは、前述の実施例と同様である。そして、文字列注目係数テーブル 3 1 0 4 に当該単語の ID があるか判定し (S 3 9 0 6)、有る場合には文字列注目係数テーブル 3 1 0 4 から、当該単語 ID に対応する文字列注目係数を読み出し (S 3 9 0 7)、その文字列注目係数を単語重要度に乘じて、算出した積を単語重要度として用いる (S 3 9 0 8)。

40

【 0 0 6 5 】

これにより、要約文書内で注目された文字列内の単語を優先した単語重要度が算出されるようになる。そして、この単語重要度を用いて文重要度を算出する (図 8)。

【 0 0 6 6 】

実施の形態 3 .

本実施の形態では、表示された要約文書中から操作者が選択した部位 (文あるいは文字列) に基づいて、当該部位の出所である要約対象文書に含まれる単語に対して重み付けを施して、要約文書を再生成する実施例について説明する。この例では、単語に対して重み

50

付けとしてBIASd（文書注目係数）を用いる。

【0067】

図40は、実施の形態3に係る全体処理フローを示す図である。本実施の形態では、要約文書提示部504による要約文書提示処理（S4004）で、文書注目係数を用いた文重要度の算出を行う。本処理については、図50を用いて後述する。文書注目係数算出部4301による文書注目係数算出処理（S4007）では、選択された部位を含む要約対象文書に基づいて文書注目係数を設定する処理を行なう。本処理については、図45を用いて後述する。

【0068】

図41は、要約文書提示処理フローを示す図である。注目文書判定処理（S4106）に続いて非注目文書判定処理（S4107）を行なう。

10

【0069】

非注目文書判定処理（S4107）について詳述する。図42は、非注目文書判定処理フローを示す図である。要約文書構成テーブル503から文書IDを順次読み出し、以下の処理を繰り返す（S4201）。当該文書IDが注目文書テーブル505に記憶されているいずれかの注目文書IDと一致するか判定し（S4202）、一致しない場合に当該文書IDを非注目文書のIDとして非注目文書テーブル4303に追加記憶させる（S4203）。但し、重複する文書IDは除く。そして、すべての文書IDについて処理した時点で終了する（S4204）。

【0070】

20

文書注目係数算出部4301による文書注目係数算出処理（S4007）について詳述する。図43は、要約文書提示及び文書注目係数算出に係るモジュール構成を示す図である。対話型複数文書要約装置は、文書注目係数算出部4301及び文書注目係数テーブル4302を有している。

【0071】

図44は、文書注目係数算出部の内部モジュール構成を示す図である。文書注目係数算出部4301は、注目文書内単語抽出部4401、注目文書内単語テーブル4402、注目文書外単語抽出部4403、注目文書外単語テーブル4404、及び係数設定部4405を有している。

【0072】

30

図45は、文書注目係数算出処理フローを示す図である。注目文書内単語抽出処理（S4501）では、操作者から指示された部位を含む文（注目文）の出所である注目文書に含まれる単語を抽出する。詳しくは、図46を用いて後述する。また、注目文書外単語抽出処理（S4502）では、それ以外の単語を抽出する。詳しくは、図47を用いて後述する。係数設定処理（S4503）では、文書注目係数を設定する。詳しくは、図48を用いて後述する。

【0073】

まず、注目文書内単語抽出処理（S4501）について詳述する。図46は、注目文書内単語抽出処理フローを示す図である。注目文書テーブル505から注目文書IDを順次読み出し、以下の処理を繰り返す（S4601）。当該注目文書IDの文書に含まれる文を順次読み出し、以下の処理を繰り返す（S4602）。文毎の処理として、当該文を解析し、当該文に含まれる単語群を特定し（S4603）、単語テーブル603から当該単語に対応する単語IDを検索し（S4604）、特定した単語IDを注目文書内単語IDとして注目文書内単語テーブル4402に記憶させる（S4605）。これらの処理をすべての単語について処理すると（S4606）、次の文の処理に移行する。そしてすべての文について処理すると（S4607）、次の注目文書IDの処理に移行し、すべての注目文書IDについて処理した時点で終了する（S4608）。

40

【0074】

次に、注目文書外単語抽出処理（S4502）について詳述する。図47は、注目文書外単語抽出処理フローを示す図である。単語テーブル602から単語IDを順次読み出し

50

、以下の処理を繰り返す（S 4 7 0 1）。当該単語IDは、注目文書内単語テーブル4 4 0 2に記憶しているいずれかの注目文書内単語IDと一致するか判定し（S 4 7 0 2）、一致しない場合に当該単語IDを注目文書外単語IDとして注目文書外単語テーブル4 4 0 4に記憶させる（S 4 7 0 3）。そして、すべての単語IDについて処理した時点で終了する（S 4 7 0 4）。

【0075】

次に、係数設定処理（S 4 5 0 3）について詳述する。図4 8は、係数設定処理フローを示す図である。注目文書内単語テーブル4 4 0 2から注目文書内単語ID群を読み出し（S 4 8 0 1）、すべての注目文書内単語IDについて、注目文書内単語IDに注目文書内単語に係る文書注目係数（1.0より大きい数値、例えば、定数2.0）を対応付けて文書注目係数テーブル4 3 0 2に記憶させる（S 4 8 0 2）。また、注目文書外単語テーブル4 4 0 4から注目文書外単語ID群を読み出し（S 4 8 0 3）、すべての注目文書外単語IDについて、注目文書外単語IDに注目文書外単語に係る文書注目係数（1.0より小さい数値、例えば、定数0.5）を対応付けて文書注目係数テーブル4 3 0 2に記憶させる（S 4 8 0 4）。

10

【0076】

続いて、要約文書生成処理（S 4 0 0 3）内の文重要度計算処理（S 7 0 2）について詳述する。図4 9は、文重要度計算に係るモジュール構成を示す図である。図5 0は、単語重要度算出処理フローを示す図である。S 5 0 0 5までは、前述の例と同様である。文書注目係数テーブルに当該単語のIDがあるか判定し（S 5 0 0 6）、ある場合には文書注目係数テーブル4 3 0 2から、当該単語IDに対応する文書注目係数を読み出し（S 5 0 0 7）、文書注目係数を単語重要度に乗じて、算出した積を単語重要度として用いる（S 5 0 0 8）。

20

【0077】

これにより、要約文書内で注目された文の出所である要約対象文書に含まれる単語を優先した単語重要度が算出されるようになる。そして、この単語重要度を用いて文重要度を算出する（図8）。

【0078】

実施の形態4.

実施の形態2の注目文字列内単語係数と実施の形態3の注目文書内単語係数を併せて用いることもできる。

30

【0079】

図5 1は、実施の形態4に係る全体処理フローを示す図である。

【0080】

要約文書提示処理（S 5 1 0 4）では、注目文字列と注目文の両方を取得する。そして、注目文字列内単語係数算出処理（S 5 1 0 7）と注目文書内単語係数算出処理（S 5 1 0 8）を行なう。

【0081】

また、要約文書生成処理（S 5 1 0 3）内の文重要度計算処理（S 7 0 2）では、両係数を単語重要度に乗算する。図5 2は、単語重要度算出処理フローを示す図である。文字列注目係数テーブル3 1 0 4に当該単語のIDがある場合には（S 5 2 0 2）、文字列注目係数テーブル3 1 0 4から、当該単語IDに対応する文字列注目係数を読み出し（S 5 2 0 3）、文字列注目係数を単語重要度に乗じて、算出した積を単語重要度として用いる（S 5 2 0 4）。更に、文書注目係数テーブル4 3 0 2に当該単語のIDがある場合には（S 5 2 0 5）、文書注目係数テーブル4 3 0 2から、当該単語IDに対応する文書注目係数を読み出し（S 5 2 0 6）、更に文書注目係数を単語重要度に乗じて、算出した積を単語重要度として用いる（S 5 2 0 7）。

40

【0082】

実施の形態5.

上述の実施例のように検索条件文を変換した検索条件式による文書検索に基づく要約文

50

書を主要約文書とし、更に検索条件文から生成した検索質問ベクトルによる文書検索に基づく推薦要約文書を生成し、両要約文書を同時に提示し、両要約文書画面から前述の注目文や注目文字列の指示の入力を受け付ける実施例について説明する。

【0083】

図53は、実施の形態5に係る全体処理フローを示す図である。

【0084】

検索条件入力処理(S5301)と(主)要約対象文書選択処理(S5302)は、前述と同様である。初期検索質問ベクトル算出部5401による初期検索質問ベクトル算出処理(S5303)では、検索条件文に基づいて検索質問ベクトルの初期化を行なう。本処理については、図57を用いて後述する。(主)要約文書生成処理(S5304)は、
10
前述と同様である。推薦要約対象文書抽出部5403による推薦要約対象文書抽出処理(S5305)では、文書ベクトルと検索質問ベクトルの類似度を算出し、類似度の大きい順に文書を抽出する処理を行なう。例えば、文書検索エンジン203は、204から順次文書ベクトルを読み出し、その文書ベクトルと検索質問ベクトルのなす角を算出し、その角の余弦値を算出して類似度とする。そして、余弦値の大きい順に所定数の文書のIDを推薦要約対象文書の文書IDとして出力する。推薦要約文書生成処理(S5306)では、
20
検索された推薦要約対象文書に基づいて推薦要約文書を生成する。本処理については、図61を用いて後述する。要約文書提示部504による要約文書同時提示処理(S5307)では、(主)要約文書画面と推薦要約文書画面を同時に出力し、いずれの要約文書からでも注目部位の指示を受け付ける。本処理については、図63を用いて後述する。(主)
20
要約対象文書絞込処理(S5309)は、前述と同様である。検索質問ベクトル更新部6401による検索質問ベクトル更新処理(S5310)では、注目文書と非注目文書に基づいて検索質問ベクトルを更新する。本処理については、図66を用いて後述する。

【0085】

図54は、初期検索質問ベクトル算出と推薦要約対象文書抽出に係るモジュール構成を示す図である。対話型複数文書要約装置は、初期検索質問ベクトル算出部5401、検索質問ベクトル記憶部5402、推薦要約対象文書抽出部5403、及び推薦要約対象文書記憶部5404を有している。

【0086】

図55は、検索質問ベクトル記憶部の構成例を示す図である。単語毎に検索質問ベクトルの成分を記憶するように構成されている。

【0087】

図56は、初期検索質問ベクトル算出部の内部モジュール構成を示す図である。初期検索質問ベクトル算出部5401は、検索条件内単語出現頻度算出部5601、検索条件内単語出現頻度テーブル5602、及び初期検索質問ベクトル成分算出部5603を有している。

【0088】

初期検索質問ベクトル算出部5401による初期検索質問ベクトル算出処理(S5303)について詳述する。図57は、初期検索質問ベクトル算出処理フローを示す図である。検索条件内単語出現頻度算出部5601による検索条件内単語出現頻度算出処理(S5701)では、検索条件内における単語の出現頻度を算出する。本処理については、図58を用いて後述する。そして、初期検索質問ベクトル成分算出部5603による初期検索質問ベクトル成分算出処理(S5702)では、検索条件内単語出現頻度に基づいて初期検索質問ベクトル成分を算出する。本処理については、図60を用いて後述する。

【0089】

検索条件内単語出現頻度算出部5601による検索条件内単語出現頻度算出処理(S5701)について詳述する。図58は、検索条件内単語出現頻度算出処理フローを示す図である。まず、検索条件内単語出現頻度テーブル5602のすべての検索条件内単語出現頻度を0にする(S5801)。そして、検索条件入力部201によるS5301の処理で受け付けた検索条件文を解析し、検索条件文に含まれる単語群を抽出し(S5802)

10

20

30

40

50

、抽出した単語毎に以下の処理を繰り返す（S5803）。単語毎の処理として、単語テーブル603から当該単語に対応する単語IDを検索し（S5804）、検索条件内単語出現頻度テーブル5602上で、当該単語IDに対応する検索条件内単語出現頻度をインクリメント（+1）する（S5805）。これをすべての単語について行なう（S5806）。

【0090】

図59は、検索条件内単語出現頻度テーブルの構成例を示す図である。単語毎に検索条件内単語出現頻度を記憶するように構成されている。

【0091】

次に、初期検索質問ベクトル成分算出部5603による初期検索質問ベクトル成分算出処理（S5702）について詳述する。図60は、初期検索質問ベクトル成分算出処理フローを示す図である。まず、検索質問ベクトル記憶部5402の検索質問ベクトルのすべての成分を0にする（S6001）。そして、検索条件内単語出現頻度テーブル5602に含まれる検索条件内単語出現頻度のうち最大値を求める（S6002）。続いて、検索条件文に含まれる単語毎に以下の処理を繰り返す（S6003）。当該単語IDに対応する検索条件内単語出現頻度を検索条件内単語出現頻度テーブル5602から読み出し（S6004）、検索条件内単語出現頻度を検索条件内単語出現頻度の最大値で割って、商を求める（S6005）。当該単語IDに対応する文書頻度逆数を文書頻度逆数テーブル606から読み出し（S6006）、前記商に1を加え、その和に文書頻度逆数を乗じて、積を求め、その積を初期検索質問ベクトルの成分として、当該単語IDに対応付けて検索質問ベクトル記憶部5402に記憶させる（S6007）。そして、すべての単語について処理した時点で終了する（S6008）。

【0092】

図61は、推薦要約文書生成と要約文書提示に係るモジュール構成を示す図である。対話型複数文書要約装置は、推薦要約文書生成部6101、推薦要約文書記憶部6102、推薦要約文書構成テーブル6103、及び推薦要約対象文書記憶部6104を有する。推薦要約文書生成の手順は、（主）要約文書生成と同様である。但し、要約の対象となる文書が異なるので別に行なわれる。

【0093】

図62は、要約文書提示部の内部モジュール構成を示す図である。要約文書提示部504は、非注目文書判定部6201を有している。

【0094】

図63は、要約文書提示処理フローを示す図である。まず、注目文記憶部2405、注目文書テーブル505、及び非注目文書テーブル4303をクリアする（S6301）。（主）要約文書画面生成処理（S6302）と推薦要約文書画面生成処理（S6303）に続いて、（主）要約文書画面出力処理（S6304）と推薦要約文書画面出力処理（S6305）を行なう。両画面は、例えば一体の枠内に上下あるは左右に並べて表示する。注目部位取得部2404による注目文取得処理（S6307）では、いずれの要約文書からでも注目文を取得する。絞込要求指示を受け付けた場合には、注目文書判定部2406による注目文書判定処理（S6308）では、推薦要約文書画面から注目文指示を受け付けた場合には、推薦要約文書構成テーブル6103に基づいて注目文書を特定する。続いて、非注目文書判定部6201による非注目文書判定処理（S6309）を行なう。この処理では、注目文書以外の非注目文書を判定する。特に、図42のS4201では、要約文書構成テーブル503及び推薦要約文書構成テーブル6103から文書IDを順次読み出し、以下の処理を繰り返す。

【0095】

図64は、検索質問ベクトル更新に係るモジュール構成を示す図である。対話型複数文書要約装置は、検索質問ベクトル更新部6401を有している。また、文書データベース204は、内部に設けた文書ベクトルテーブル6802から文書ベクトルを提供する。

【0096】

10

20

30

40

50

図65は、文書ベクトルテーブルの構成例を示す図である。文書毎に文書ベクトルを記憶するように構成されており、文書ベクトルは、各単語に対応する文書ベクトル成分から構成されている。文書ベクトルの算出については、図70を用いて後述する。

【0097】

検索質問ベクトル更新部6401による検索質問ベクトル更新処理(S5310)について詳述する。図66は、検索質問ベクトル更新処理フローを示す図である。まず、加算修正ベクトル算出処理(S6601)で、検索質問ベクトルを修正するために加算するベクトル値を算出する。本処理については、図67を用いて後述する。次に、減算修正ベクトル算出処理(S6602)で、検索質問ベクトルを修正するために減算するベクトル値を算出する。本処理については、図68を用いて後述する。そして、検索質問ベクトル記憶部から旧検索質問ベクトルを読み出し(S6603)、旧検索質問ベクトルに加算修正ベクトルを加算し、更に減算修正ベクトルを減算し、新検索質問ベクトルを求める(S6604)。求めた新検索質問ベクトルを検索質問ベクトル記憶部5402に書き込む(S6605)。

10

【0098】

図67は、加算修正ベクトル算出処理フローを示す図である。注目文書テーブル505に記憶している各注目文書IDについて、以下の処理を繰り返す(S6701)。当該注目文書IDに対応する文書ベクトルを読み出し(S6702)、文書ベクトルを注目文書累計ベクトル(初期値:0ベクトル)に加算する(S6703)。すべての注目文書IDについて処理すると(S6704)、注目文書累計ベクトルを注目文書の総数で割って、注目文書平均ベクトルを求め(S6705)、注目文書平均ベクトルに注目文書更新寄与係数(例えば、2.0)を乗じて、加算修正ベクトルを得る(S6706)。

20

【0099】

図68は、減算修正ベクトル算出処理フローを示す図である。非注目文書テーブル4303に記憶している各非注目文書IDについて、以下の処理を繰り返す(S6801)。当該非注目文書IDに対応する文書ベクトルを読み出し(S6802)、文書ベクトルを非注目文書累計ベクトル(初期値:0ベクトル)に加算する(S6803)。すべての非注目文書IDについて処理すると(S6804)、非注目文書累計ベクトルを非注目文書の総数で割って、非注目文書平均ベクトルを求め(S6805)、非注目文書平均ベクトルに非注目文書更新寄与係数(例えば、0.5)を乗じて、減算修正ベクトルを得る(S6806)。

30

【0100】

ここで、文書データベース204内部における文書ベクトル算出について説明する。図68は、文書ベクトル算出に係る文書データベースのモジュール構成を示す図である。文書データベース204は、文書ベクトル算出部6901を有している。

【0101】

図70は、文書ベクトル算出処理フローを示す図である。文書管理テーブル602で管理している各文書について、以下の処理を繰り返す(S7001)。文書管理テーブル602から当該文書の文書内単語総数を読み出す(S7002)。そして、各単語について、以下の処理を繰り返す(S7003)。文書内単語出現頻度テーブル605から当該文書の当該単語に係る文書内単語出現頻度を読み出し(S7004)、更に文書頻度逆数テーブル606から当該単語の文書頻度逆数を読み出す(S7005)。そして、文書内単語出現頻度を文書内単語総数で割り、その商に文書頻度逆数を乗じて積を求め(S7006)、算出した積を当該文書の当該単語に係る文書ベクトル成分として、文書ベクトルテーブル6902に記憶させる(S7007)。すべての単語について処理すると(S7008)、次の文書の処理に移行し、すべての文書について処理した時点で終了する(S7009)。これにより、図65に示した文書ベクトルテーブル6902が得られる。

40

【0102】

実施の形態6.

実施の形態5に実施の形態2の特徴を加えることも有効である。図71は、実施の形態

50

6に係る全体処理フローを示す図である。

【0103】

実施の形態7.

実施の形態5に実施の形態3の特徴を加えることも有効である。図72は、実施の形態7に係る全体処理フローを示す図である。

【0104】

実施の形態8.

実施の形態5に実施の形態4の特徴を加えることも有効である。図73は、実施の形態8に係る全体処理フローを示す図である。

【0105】

実施の形態9.

前述の例では、主要約文書画面と推薦要約文書画面を同時に表示したが、一方のみを表示し、切替指示を受け付けた場合に他方の画面に切り換えることもできる。図74は、実施の形態9に係る全体処理フローを示す図である。

【0106】

要約文書提示部504による要約文書(切替)提示処理(S7407)では、両画面の切り換えを行なう。図75と図76は、要約文書(切替)提示処理フローを示す図である。(主)要約画面を出力中に画面の切替指示を受け付けた場合には(S7504)、推薦要約文書画面生成処理(S7511)に移行し、推薦要約文書画面出力処理(S7512)を行なう。また、推薦要約画面を出力中に画面の切替指示を受け付けた場合には(S7513)、(主)要約文書画面生成処理(S7502)に移行し、(主)要約文書画面出力処理(S7503)を行なう。S7507以下では、要約文書構成テーブル503を用いて注目文書と非注目文書を判定し、S7516以下では、推薦要約文書構成テーブル6103を用いて注目文書と非注目文書を判定する。

【0107】

実施の形態10.

実施の形態2から要約対象文書絞込の処理を除くこともできる。この場合には、同じ要約対象文書群に基づいて、新たな要約文書が生成される。図77は、実施の形態10に係る全体処理フローを示す図である。

【0108】

実施の形態11.

実施の形態3から要約対象文書絞込の処理を除くこともできる。この場合には、同じ要約対象文書群に基づいて、新たな要約文書が生成される。図78は、実施の形態11に係る全体処理フローを示す図である。

【0109】

実施の形態12.

実施の形態4から要約対象文書絞込の処理を除くこともできる。この場合には、同じ要約対象文書群に基づいて、新たな要約文書が生成される。図79は、実施の形態12に係る全体処理フローを示す図である。

【0110】

実施の形態13.

実施の形態5から主要約文書に関する処理を除くこともできる。この場合には、推薦要約文書のみが生成され、提示される。図80は、実施の形態13に係る全体処理フローを示す図である。

【0111】

実施の形態14.

実施の形態6から主要約文書に関する処理を除くこともできる。この場合には、推薦要約文書のみが生成され、提示される。図81は、実施の形態14に係る全体処理フローを示す図である。

【0112】

10

20

30

40

50

実施の形態 15 .

実施の形態 7 から主要約文書に関する処理を除くこともできる。この場合には、推薦要約文書のみが生成され、提示される。図 8 2 は、実施の形態 15 に係る全体処理フローを示す図である。

【 0 1 1 3 】

実施の形態 16 .

実施の形態 8 から主要約文書に関する処理を除くこともできる。この場合には、推薦要約文書のみが生成され、提示される。図 8 3 は、実施の形態 16 に係る全体処理フローを示す図である。

【 0 1 1 4 】

実施の形態 17 .

本実施の形態では、検索条件文に含まれる単語に対して重み付けを施して、要約文書を再生成する実施例について説明する。この例では、単語に対して重み付けとして B I A S s (検索キーワード係数) を用いる。

【 0 1 1 5 】

図 8 4 は、実施の形態 17 に係る全体処理フローを示す図である。検索キーワード係数算出部 8 5 0 1 による検索キーワード係数算出処理 (8 3 0 3) で算出した検索キーワード係数を用いて、文重要度を計算する。

【 0 1 1 6 】

図 8 5 は、検索キーワード係数算出に係るモジュール構成を示す図である。対話型複数文書要約装置は、検索キーワード係数算出部 8 5 0 1 及び検索キーワード係数テーブル 8 5 0 2 を有している。

【 0 1 1 7 】

図 8 6 は、検索キーワード係数算出処理フローを示す図である。検索条件文を解析し、検索条件文に含まれる単語として検索キーワード群を特定する (S 8 6 0 1)。そして、検索キーワード毎に、以下の処理を繰り返す (S 8 6 0 2)、当該検索キーワードの単語に対応する単語 ID を特定し (S 8 6 0 3)、単語 ID に検索キーワード係数 (1 . 0 より大きい数値、例えば、定数 2 . 0) を対応付けて検索キーワード係数テーブルに記憶させる (S 8 6 0 4)。すべての検索キーワードについて処理した時点で終了する (S 8 6 0 5)。

【 0 1 1 8 】

要約文書生成処理 (S 8 4 0 4) 内の単語重要度算出処理 (S 7 0 2) について詳述する。図 8 7 は、単語重要度算出処理フローを示す図である。検索キーワード係数テーブル 8 5 0 2 に当該単語の ID がある場合には (S 8 7 0 2)、検索キーワード係数テーブル 8 5 0 2 から、当該単語 ID に対応する検索キーワード係数を読み出す (S 8 7 0 3)。そして、検索キーワード係数を単語重要度に乘じて、算出した積を単語重要度として用いる (S 8 7 0 4)。

【 0 1 1 9 】

検索キーワード係数を用いる形態は、上述の実施の形態 1 乃至 16 のいずれにも適用することができる。

【 0 1 2 0 】

対話型複数文書要約装置は、コンピュータであり、各要素はプログラムにより処理を実行することができる。また、プログラムを記憶媒体に記憶させ、記憶媒体からコンピュータに読み取られるようにすることができる。

【 図面の簡単な説明 】

【 0 1 2 1 】

【 図 1 】 実施の形態 1 に係る全体処理フローを示す図である。

【 図 2 】 要約対象文書選択に係るモジュール構成を示す図である。

【 図 3 】 要約対象文書選択処理フローを示す図である。

【 図 4 】 要約対象文書記憶部の構成例を示す図である。

10

20

30

40

50

- 【図5】要約文書生成、要約文書提示、及び要約対象文書絞込に係るモジュール構成を示す図である。
- 【図6】要約文書生成部の内部モジュール構成を示す図である。
- 【図7】要約文書生成処理フローを示す図である。
- 【図8】文重要度計算処理フロー（1 / 2）を示す図である。
- 【図9】文重要度計算処理フロー（2 / 2）を示す図である。
- 【図10】文書管理テーブルの構成例を示す図である。
- 【図11】単語テーブルの構成例を示す図である。
- 【図12】文中総単語数算出処理フローを示す図である。
- 【図13】文内単語出現頻度テーブルの構成例を示す図である。 10
- 【図14】単語重要度算出処理フローを示す図である。
- 【図15】文書内単語出現頻度テーブルの構成例を示す図である。
- 【図16】文書頻度逆数テーブルの構成例を示す図である。
- 【図17】文重要度テーブルの構成例を示す図である。
- 【図18】重要文抽出処理フロー（1 / 2）を示す図である。
- 【図19】重要文抽出処理フロー（2 / 2）を示す図である。
- 【図20】文抽出テーブルの構成例を示す図である。
- 【図21】重要文整列処理フローを示す図である。
- 【図22】要約文書記憶部の構成例を示す図である。
- 【図23】要約文書構成テーブルの構成例を示す図である。 20
- 【図24】要約文書提示部の内部モジュール構成を示す図である。
- 【図25】要約文書提示処理フローを示す図である。
- 【図26】要約文書画面生成処理フローを示す図である。
- 【図27】注目文取得処理フローを示す図である。
- 【図28】注目文書判定処理フローを示す図である。
- 【図29】要約対象文書絞込処理フローを示す図である。
- 【図30】実施の形態2に係る全体処理フローを示す図である。
- 【図31】要約文書提示及び文字列注目係数算出に係るモジュール構成を示す図である。
- 【図32】要約文書提示部の内部モジュール構成を示す図である。
- 【図33】要約文書提示処理フローを示す図である。 30
- 【図34】注目文字列取得処理フローを示す図である。
- 【図35】注目文字列内単語判定処理フローを示す図である。
- 【図36】注目文字列外単語判定処理フローを示す図である。
- 【図37】文字列注目係数算出処理フローを示す図である。
- 【図38】文重要度計算に係るモジュール構成を示す図である。
- 【図39】単語重要度算出処理フローを示す図である。
- 【図40】実施の形態3に係る全体処理フローを示す図である。
- 【図41】要約文書提示処理フローを示す図である。
- 【図42】非注目文書判定処理フローを示す図である。
- 【図43】要約文書提示及び文書注目係数算出に係るモジュール構成を示す図である。 40
- 【図44】文書注目係数算出部の内部モジュール構成を示す図である。
- 【図45】文書注目係数算出処理フローを示す図である。
- 【図46】注目文書内単語抽出処理フローを示す図である。
- 【図47】注目文書外単語抽出処理フローを示す図である。
- 【図48】係数設定処理フローを示す図である。
- 【図49】文重要度計算に係るモジュール構成を示す図である。
- 【図50】単語重要度算出処理フローを示す図である。
- 【図51】実施の形態4に係る全体処理フローを示す図である。
- 【図52】単語重要度算出処理フローを示す図である。
- 【図53】実施の形態5に係る全体処理フローを示す図である。 50

【図54】初期検索質問ベクトル算出と推薦要約対象文書抽出に係るモジュール構成を示す図である。

【図55】検索質問ベクトル記憶部の構成例を示す図である。

【図56】初期検索質問ベクトル算出部の内部モジュール構成を示す図である。

【図57】初期検索質問ベクトル算出処理フローを示す図である。

【図58】検索条件内単語出現頻度算出処理フローを示す図である。

【図59】検索条件内単語出現頻度テーブルの構成例を示す図である。

【図60】初期検索質問ベクトル成分算出処理フローを示す図である。

【図61】推薦要約文書生成と要約文書提示に係るモジュール構成を示す図である。

【図62】要約文書提示部の内部モジュール構成を示す図である。

10

【図63】要約文書提示処理フローを示す図である。

【図64】検索質問ベクトル更新に係るモジュール構成を示す図である。

【図65】文書ベクトルテーブルの構成例を示す図である。

【図66】検索質問ベクトル更新処理フローを示す図である。

【図67】加算修正ベクトル算出処理フローを示す図である。

【図68】減算修正ベクトル算出処理フローを示す図である。

【図69】文書ベクトル算出に係る文書データベースのモジュール構成を示す図である。

【図70】文書ベクトル算出処理フローを示す図である。

【図71】実施の形態6に係る全体処理フローを示す図である。

【図72】実施の形態7に係る全体処理フローを示す図である。

20

【図73】実施の形態8に係る全体処理フローを示す図である。

【図74】実施の形態9に係る全体処理フローを示す図である。

【図75】要約文書(切換)提示処理フロー(1/2)を示す図である。

【図76】要約文書(切換)提示処理フロー(2/2)を示す図である。

【図77】実施の形態10に係る全体処理フローを示す図である。

【図78】実施の形態11に係る全体処理フローを示す図である。

【図79】実施の形態12に係る全体処理フローを示す図である。

【図80】実施の形態13に係る全体処理フローを示す図である。

【図81】実施の形態14に係る全体処理フローを示す図である。

【図82】実施の形態15に係る全体処理フローを示す図である。

30

【図83】実施の形態16に係る全体処理フローを示す図である。

【図84】実施の形態17に係る全体処理フローを示す図である。

【図85】検索キーワード係数算出に係るモジュール構成を示す図である。

【図86】検索キーワード係数算出処理フローを示す図である。

【図87】単語重要度算出処理フローを示す図である。

【符号の説明】

【0122】

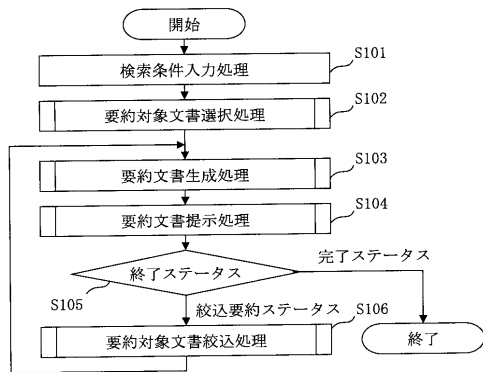
201 検索条件入力部、202 要約対象文書選択部、203 文書検索エンジン、
 204 文書データベース、205 要約対象文書記憶部、501 要約文書生成部、5
 02 要約文書記憶部、503 要約文書構成テーブル、504 要約文書提示部、50
 5 注目文書テーブル、506 要約対象文書絞込部、601 文重要度計算部、602
 文書管理テーブル、603 単語テーブル、604 文内単語出現頻度テーブル、60
 5 文書内単語出現頻度テーブル、606 文書頻度逆数テーブル、607 文提供部、
 608 文書データ記憶部、609 文重要度テーブル、610 重要文抽出部、611
 文抽出テーブル、612 重要文整列部、2401 要約文書画面生成部、2402
 要約文書画面記憶部、2403 要約文書画面出力部、2404 注目部位取得部、24
 05 注目文記憶部、2406 注目文書判定部、3101 注目文字列内単語テーブル
 、3102 注目文字列外単語テーブル、3103 文字列注目係数算出部、3104
 文字列注目係数テーブル、3201 注目文字列記憶部、3202 注目文字列内単語判
 定部、3203 注目文字列外単語判定部、4301 文書注目係数算出部、4302

40

50

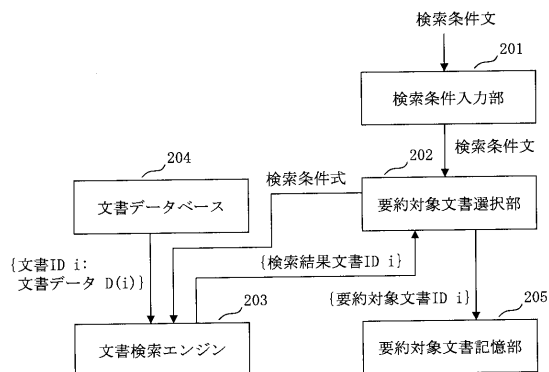
文書注目係数テーブル、4303 非注目文書テーブル、4401 注目文書内単語抽出部、4402 注目文書内単語テーブル、4403 注目文書外単語抽出部、4404 注目文書外単語テーブル、4405 係数設定部、5401 初期検索質問ベクトル算出部、5402 検索質問ベクトル記憶部、5403 推薦要約対象文書抽出部、5404 推薦要約対象文書記憶部、5601 検索条件内単語出現頻度算出部、5602 検索条件内単語出現頻度テーブル、5603 初期検索質問ベクトル成分算出部、6101 推薦要約文書生成部、6102 推薦要約文書記憶部、6103 推薦要約文書構成テーブル、6104 推薦要約対象文書記憶部、6201 非注目文書判定部、6401 検索質問ベクトル更新部、6901 文書ベクトル算出部、8501 検索キーワード係数算出部、8502 検索キーワード係数テーブル。

【図1】

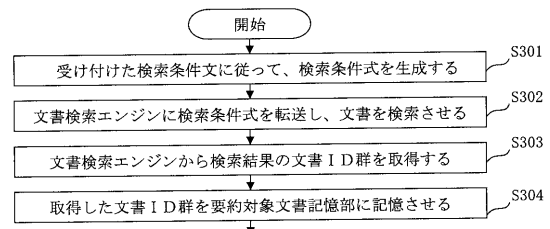


<全体処理フロー>

【図2】

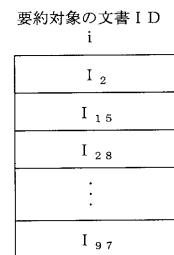


【図3】



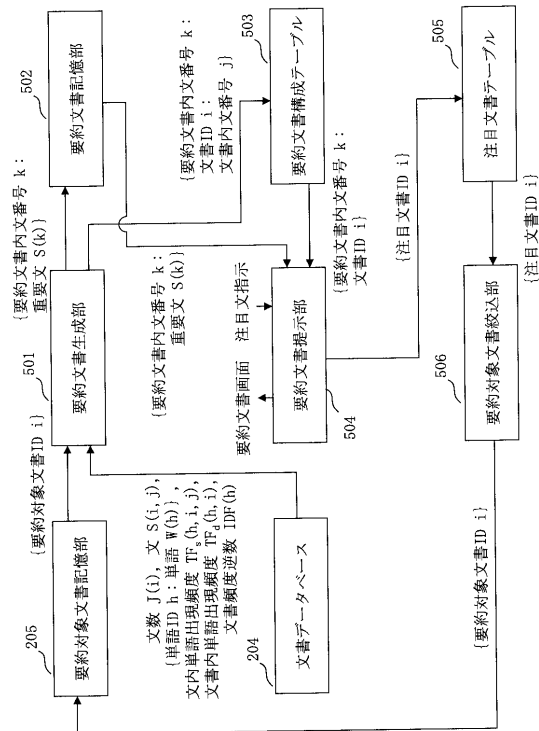
<要約対象文書選択処理>

【図4】

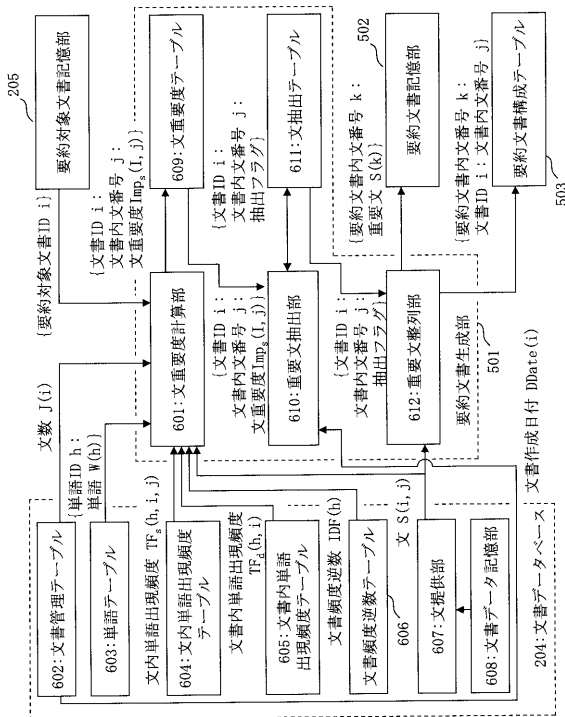


<要約対象文書記憶部>

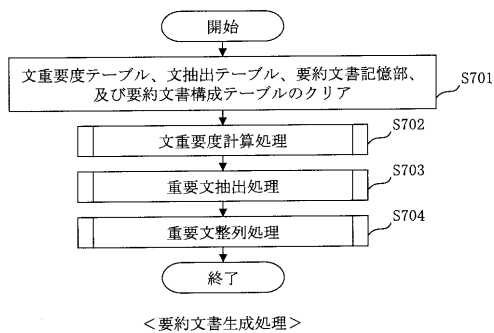
【図5】



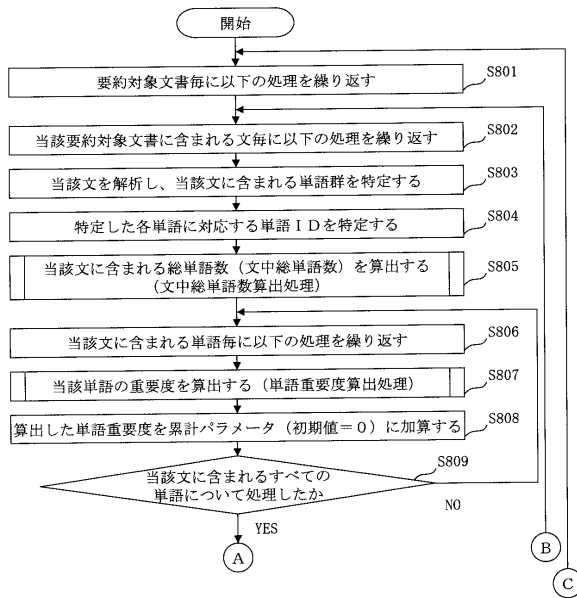
【図6】



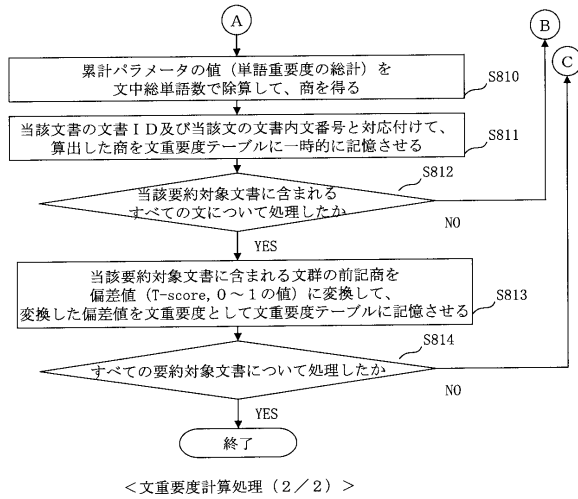
【図7】



【図8】



【図9】



【図10】

文書ID i	文書作成日付 DData(i)	文書データ ファイル名 D(I ₁).TXT	文数 J(i)	文書内単語総数 Nw(i)
I ₁	DData(I ₁)	D(I ₁).TXT	J(I ₁)	Nw(I ₁)
I ₂	DData(I ₂)	D(I ₂).TXT	J(I ₂)	Nw(I ₂)
I ₃	DData(I ₃)	D(I ₃).TXT	J(I ₃)	Nw(I ₃)
⋮	⋮	⋮	⋮	⋮
I _M	DData(I _M)	D(I _M).TXT	J(I _M)	Nw(I _M)

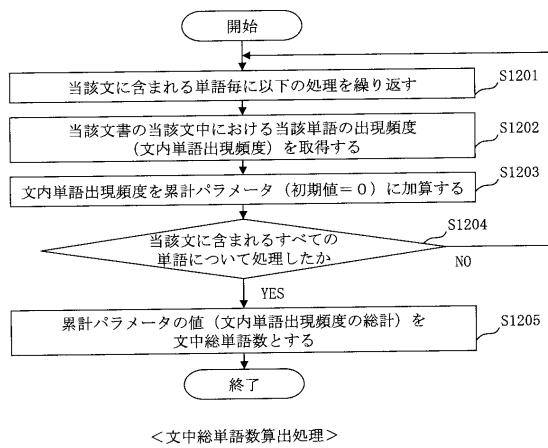
<文書管理テーブル>

【図11】

単語ID h	単語文字列 W(h)
H ₁	W(H ₁)
H ₂	W(H ₂)
H ₃	W(H ₃)
⋮	⋮
H _N	W(H _N)

<単語テーブル>

【図12】

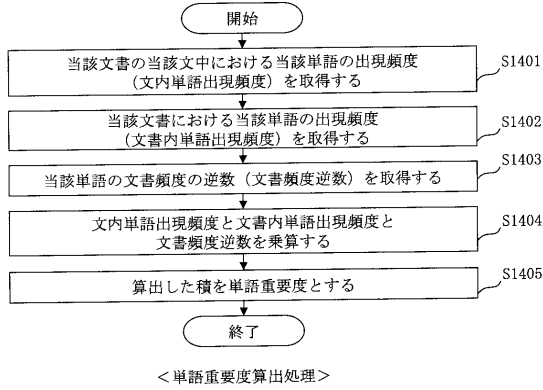


【図13】

単語ID h	文書内文番号		TF _s (H ₁ , I ₁ , J(I ₁))
	1	2	
H ₁	TF _s (H ₁ , I ₁ , 1)	TF _s (H ₁ , I ₁ , 2)	TF _s (H ₁ , I ₁ , J(I ₁))
H ₂	TF _s (H ₂ , I ₁ , 1)	TF _s (H ₂ , I ₁ , 2)	TF _s (H ₂ , I ₁ , J(I ₁))
H ₃	TF _s (H ₃ , I ₁ , 1)	TF _s (H ₃ , I ₁ , 2)	TF _s (H ₃ , I ₁ , J(I ₁))
⋮	⋮	⋮	⋮
H _N	TF _s (H _N , I ₁ , 1)	TF _s (H _N , I ₁ , 2)	TF _s (H _N , I ₁ , J(I ₁))

<文内単語出現頻度テーブル>

【図14】



【図16】

単語ID h	文書頻度逆数 IDF(h)
H ₁	IDF(H ₁)
H ₂	IDF(H ₂)
H ₃	IDF(H ₃)
⋮	⋮
H _N	IDF(H _N)

<文書頻度逆数テーブル>

【図15】

単語ID h	文書ID i			
	I ₁	I ₂	⋯	I _M
H ₁	TF _d (H ₁ , I ₁)	TF _d (H ₁ , I ₂)	⋯	TF _d (H ₁ , I _M)
H ₂	TF _d (H ₂ , I ₁)	TF _d (H ₂ , I ₂)	⋯	TF _d (H ₂ , I _M)
H ₃	TF _d (H ₃ , I ₁)	TF _d (H ₃ , I ₂)	⋯	TF _d (H ₃ , I _M)
⋮	⋮	⋮	⋮	⋮
H _N	TF _d (H _N , I ₁)	TF _d (H _N , I ₂)	⋯	TF _d (H _N , I _M)

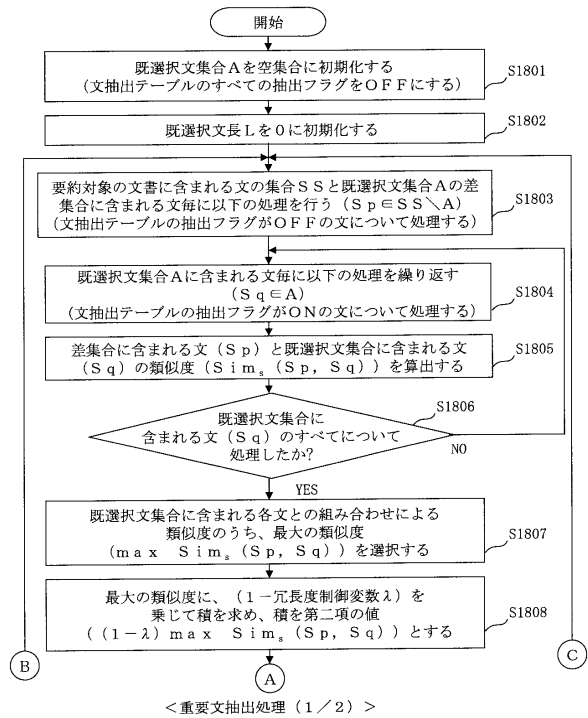
<文書内単語出現頻度テーブル>

【図17】

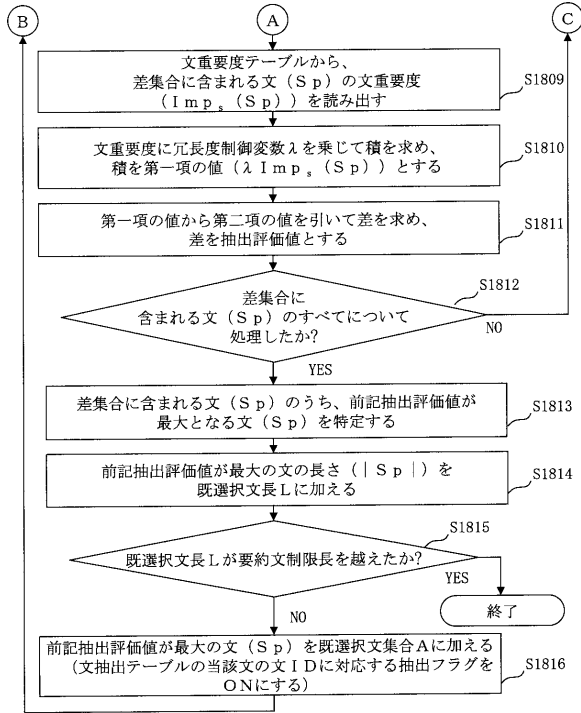
文書ID i	文書内文番号 j	文書重要度 Imp _s (i, j)
I ₂	1	Imp _s (I ₂ , 1)
I ₂	2	Imp _s (I ₂ , 2)
I ₂	3	Imp _s (I ₂ , 3)
⋮	⋮	⋮
I ₂	J(I ₂)	Imp _s (I ₂ , J(I ₂))
I ₁₅	1	Imp _s (I ₁₅ , 1)
⋮	⋮	⋮
I ₁₅	J(I ₁₅)	Imp _s (I ₁₅ , J(I ₁₅))
I ₂₈	1	Imp _s (I ₂₈ , 1)
⋮	⋮	⋮
I ₂₈	J(I ₂₈)	Imp _s (I ₂₈ , J(I ₂₈))
⋮	⋮	⋮
I ₉₇	1	Imp _s (I ₉₇ , 1)
⋮	⋮	⋮
I ₉₇	J(I ₉₇)	Imp _s (I ₉₇ , J(I ₉₇))

<文重要度テーブル>

【図18】



【図 19】



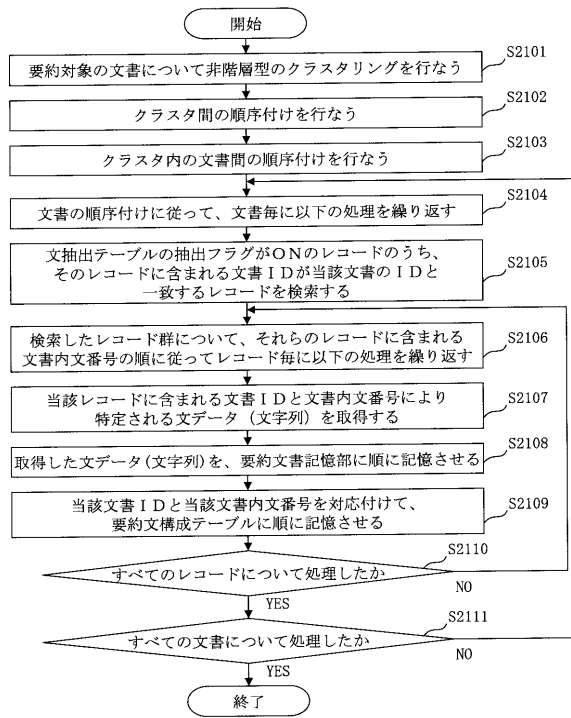
<重要文抽出処理 (2/2)>

【図 20】

文書 ID i	文書内文番号 j	抽出フラグ
I ₂	1	OFF
I ₂	2	OFF
I ₂	3	OFF
⋮	⋮	⋮
I ₂	J (I ₂)	OFF
I ₁₅	1	OFF
⋮	⋮	⋮
I ₁₅	J (I ₁₅)	OFF
I ₂₈	1	OFF
⋮	⋮	⋮
I ₂₈	J (I ₂₈)	OFF
⋮	⋮	⋮
I ₉₇	1	OFF
⋮	⋮	⋮
I ₉₇	J (I ₉₇)	OFF

<文抽出テーブル>

【図 21】



<重要文整列処理>

【図 22】

要約内文番号 k	重要文 S (k)
K ₁	S (K ₁)
K ₂	S (K ₂)
K ₃	S (K ₃)
K ₄	S (K ₄)
⋮	⋮

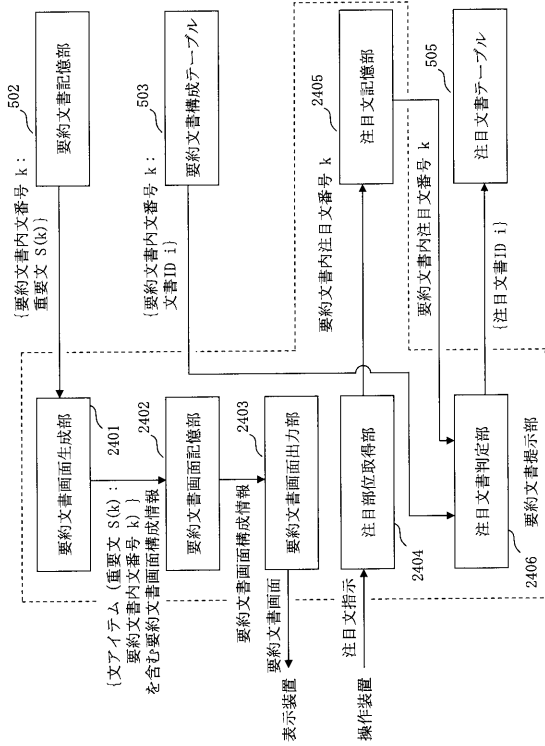
<要約文書記憶部>

【図 23】

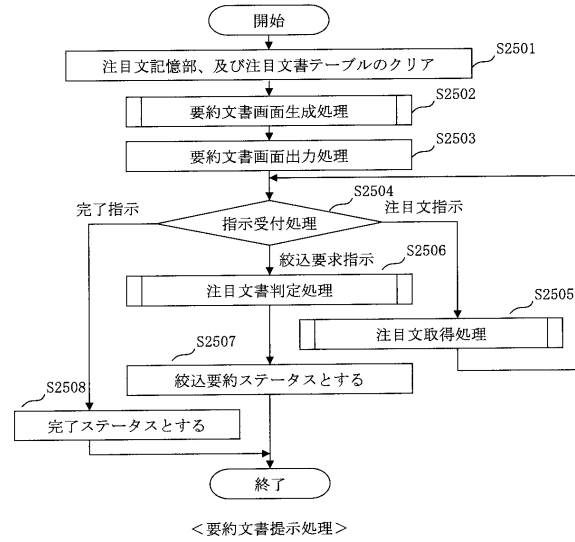
要約内文番号 k	文書 ID i	文書内文番号 j
K ₁	I ₂	8
K ₂	I ₂	14
K ₃	I ₂₈	3
K ₄	I ₃₆	26
⋮	⋮	⋮

<要約文構成テーブル>

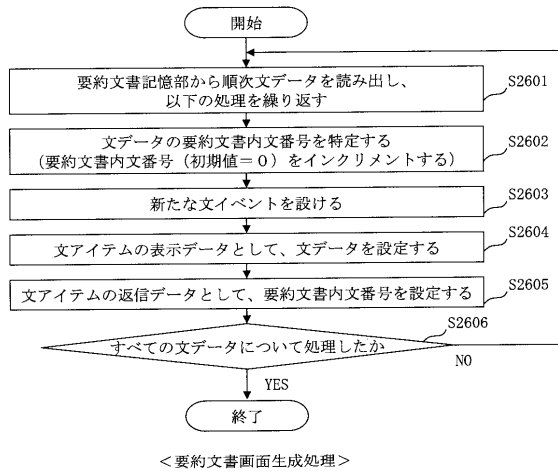
【図 24】



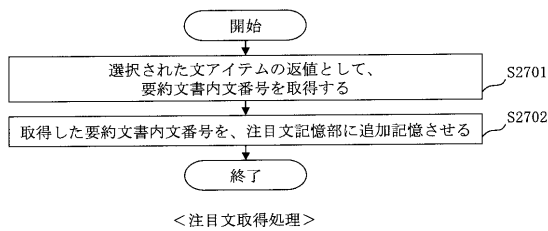
【図 25】



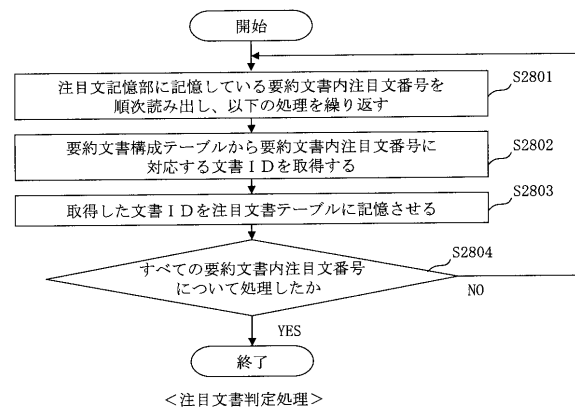
【図 26】



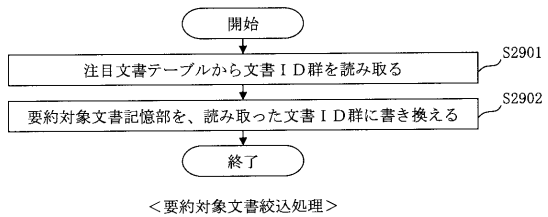
【図 27】



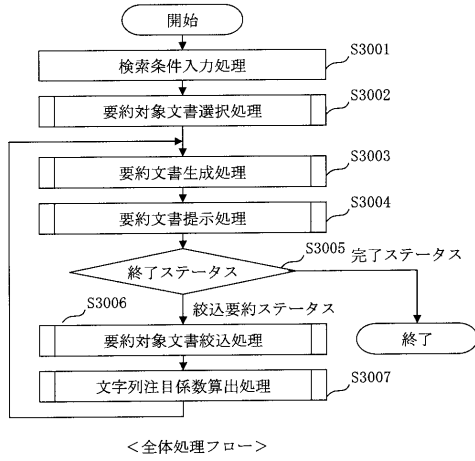
【図 28】



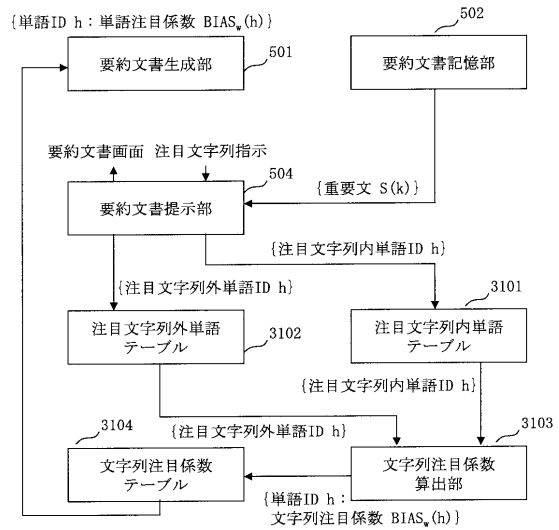
【図 29】



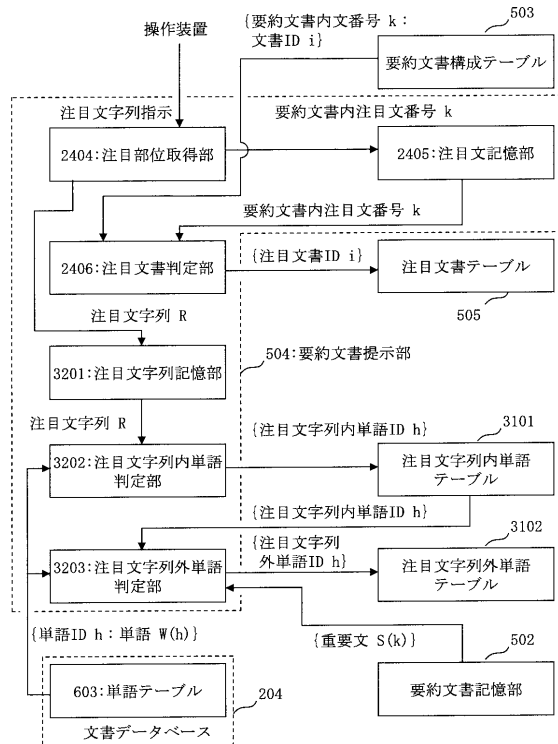
【図30】



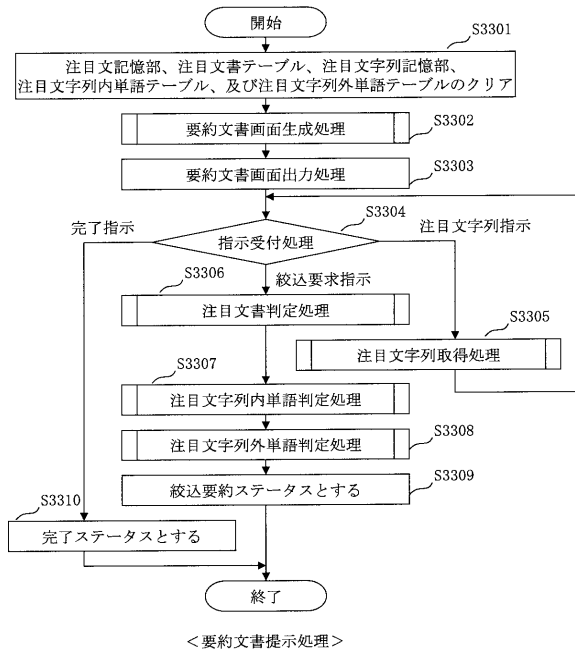
【図31】



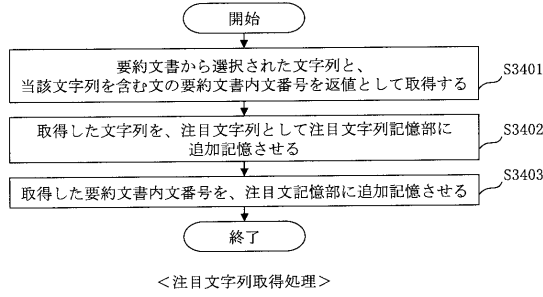
【図32】



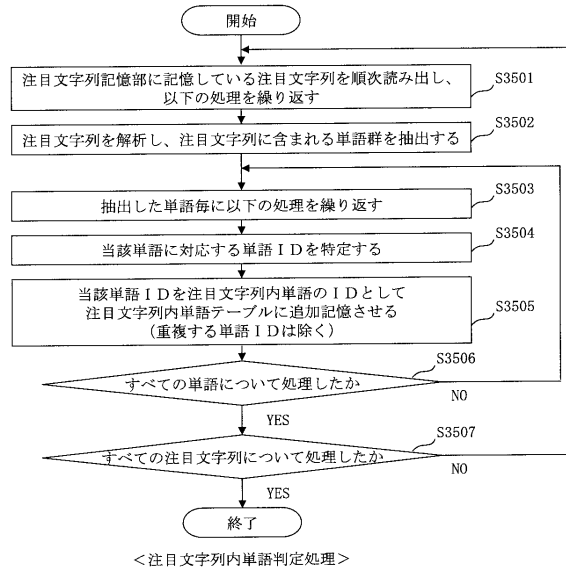
【図33】



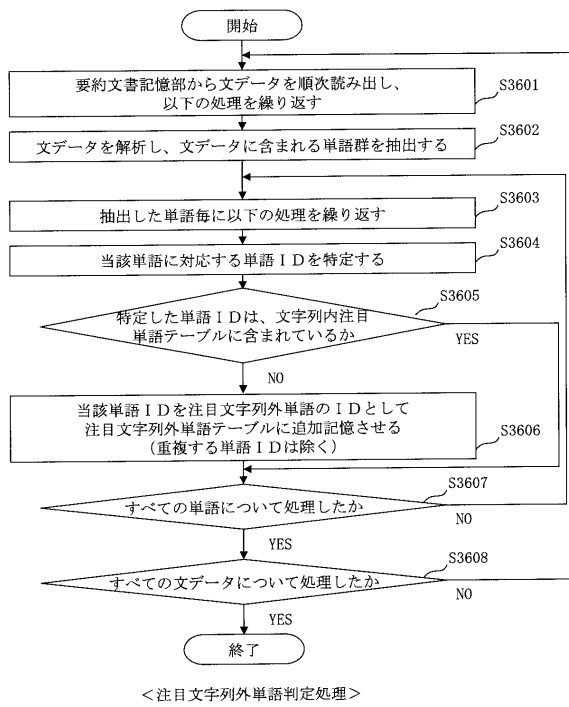
【図 3 4】



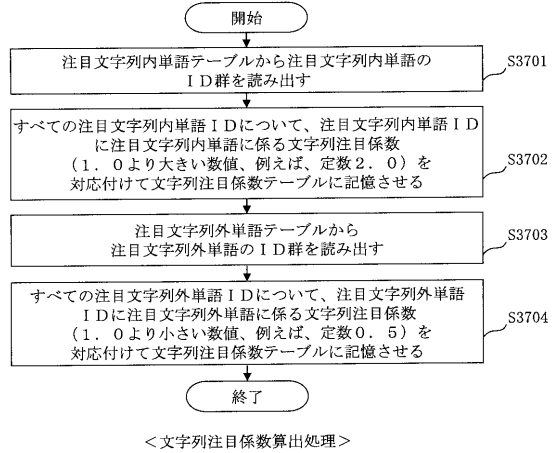
【図 3 5】



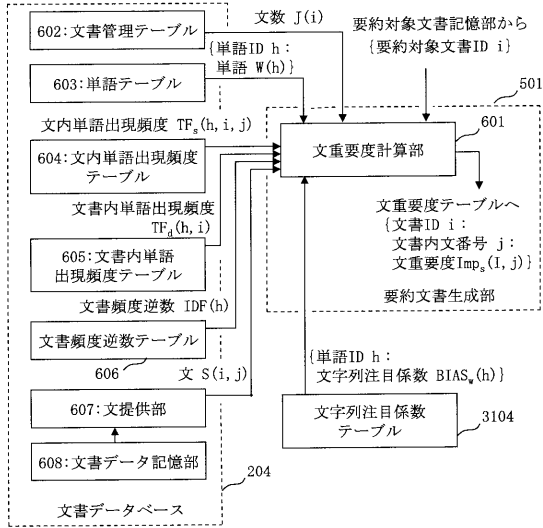
【図 3 6】



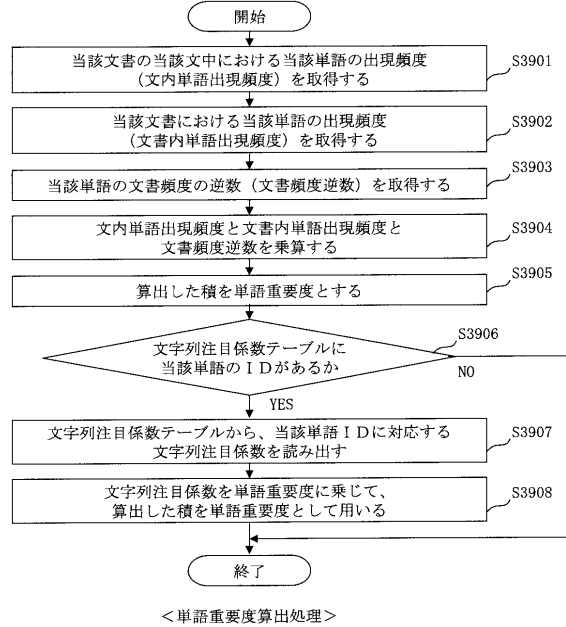
【図 3 7】



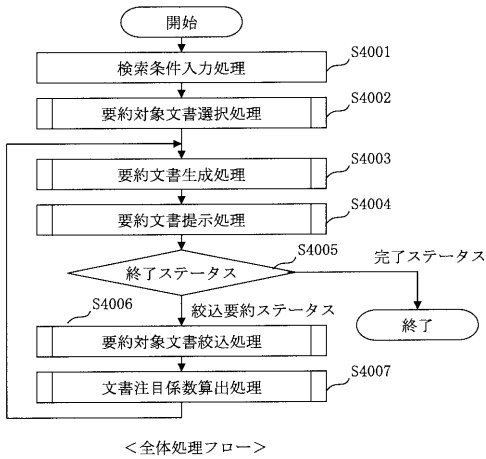
【図38】



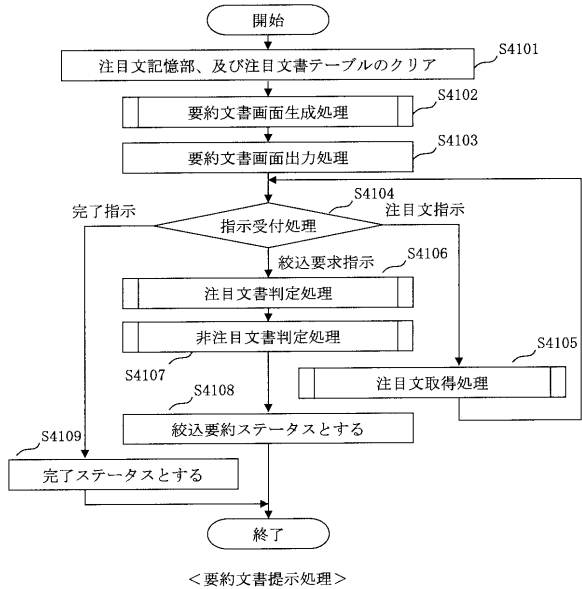
【図39】



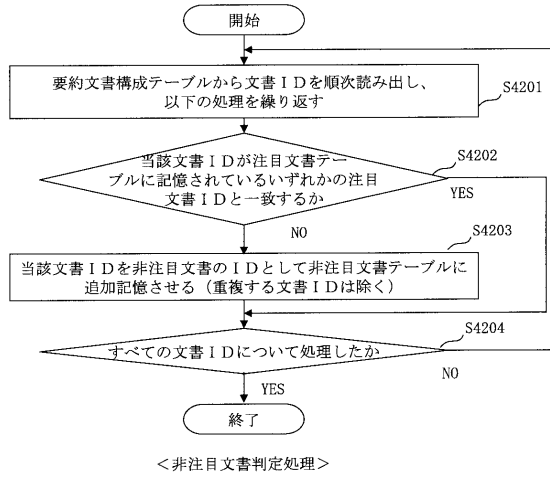
【図40】



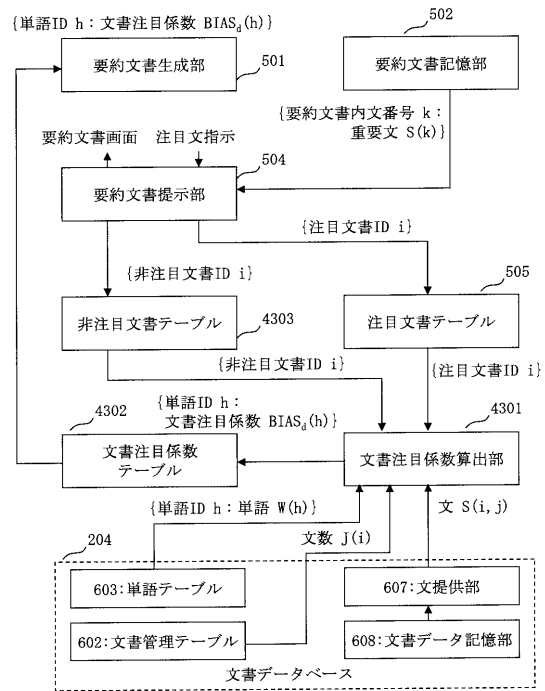
【図41】



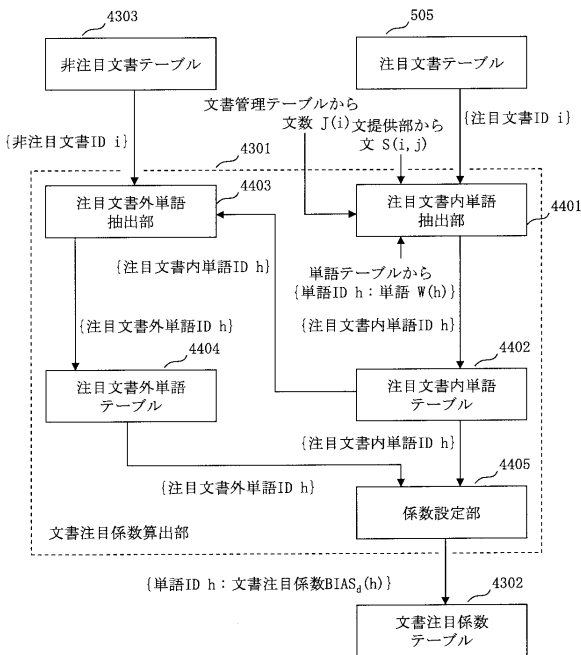
【図42】



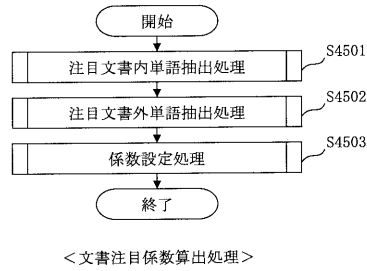
【図43】



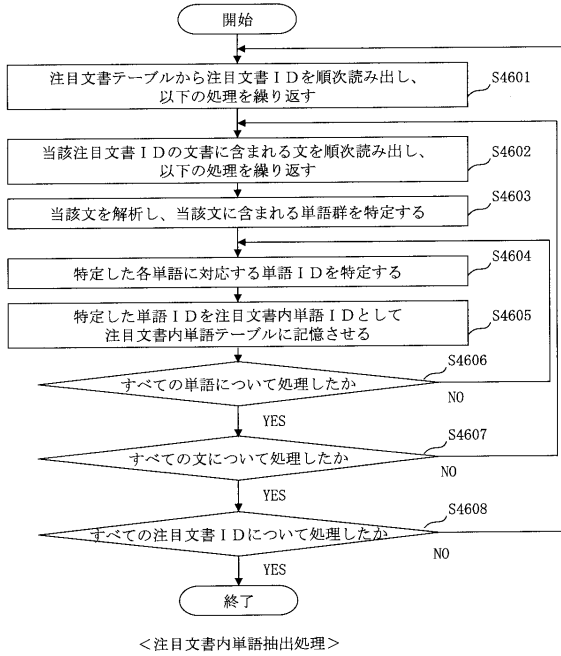
【図44】



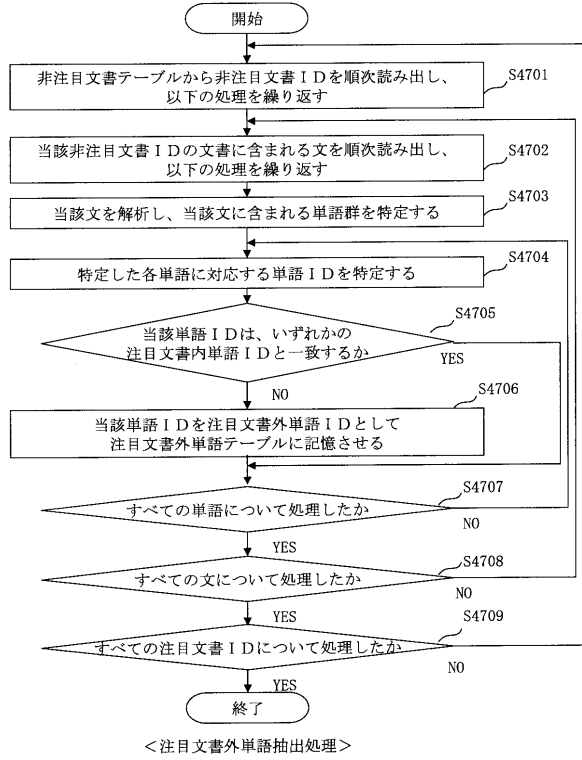
【図45】



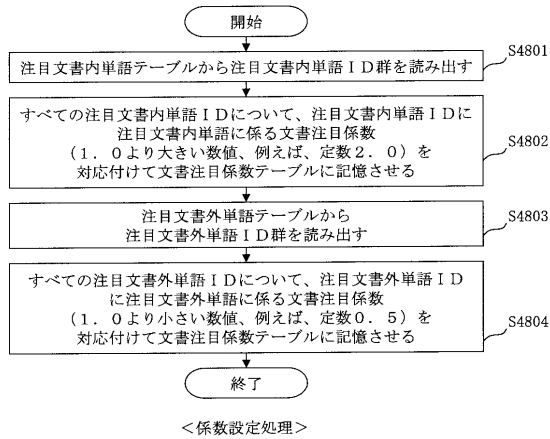
【図46】



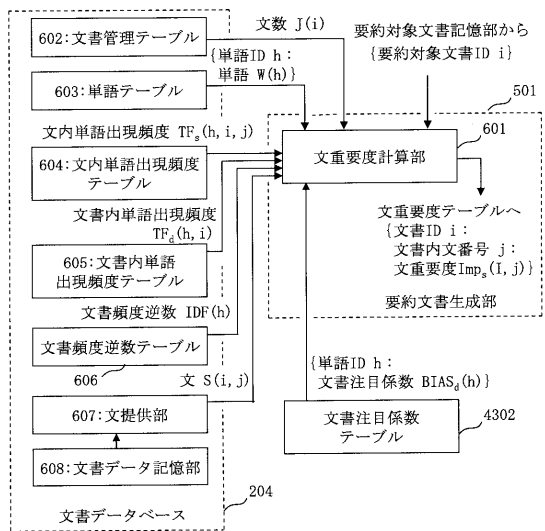
【図47】



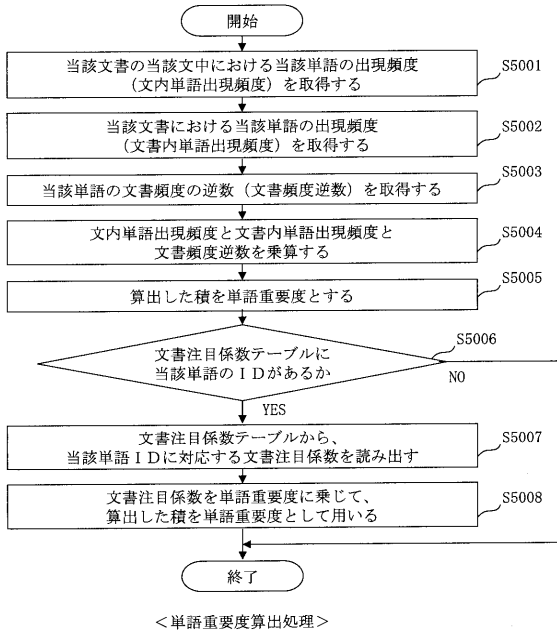
【図48】



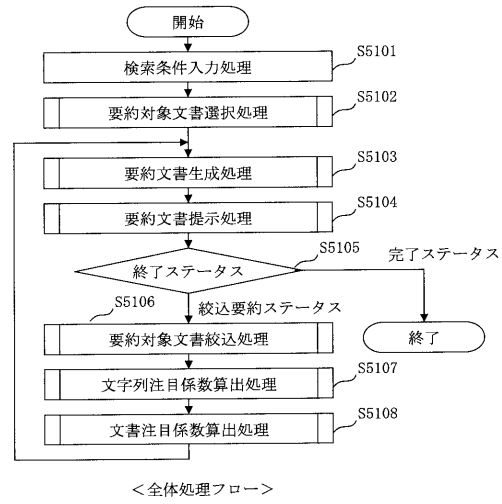
【図49】



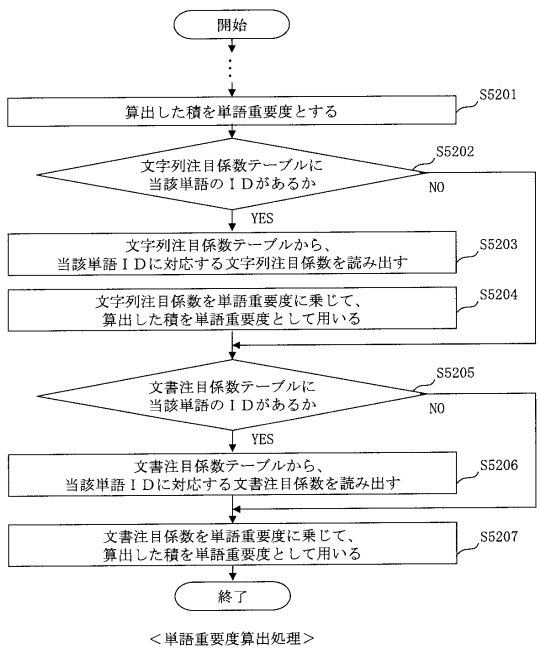
【図50】



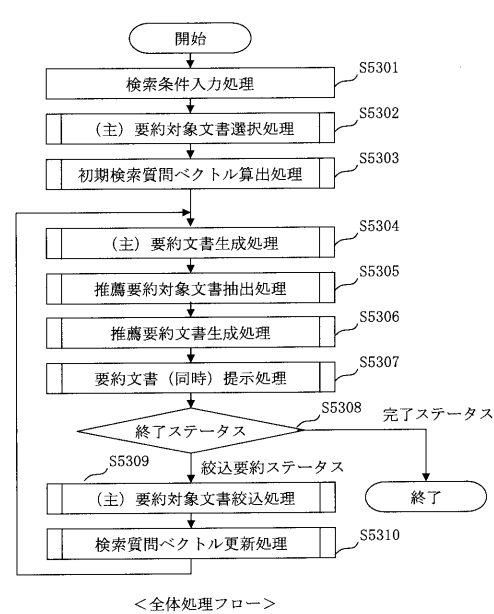
【図51】



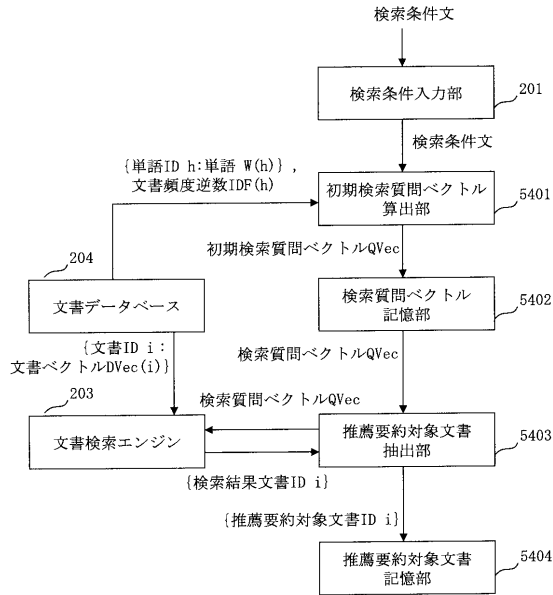
【図52】



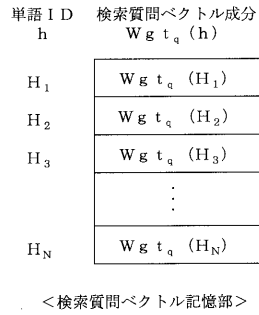
【図53】



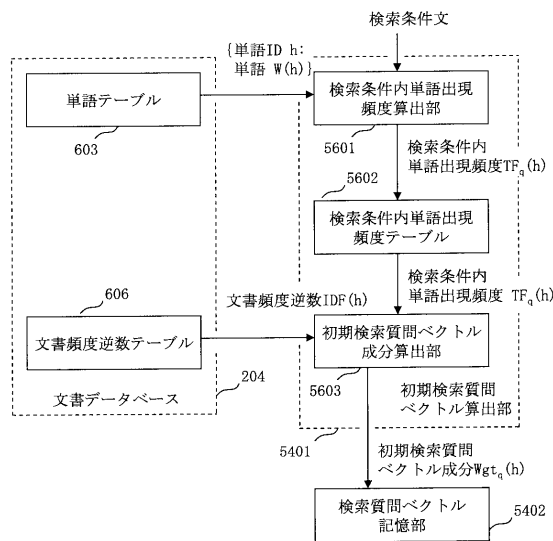
【図54】



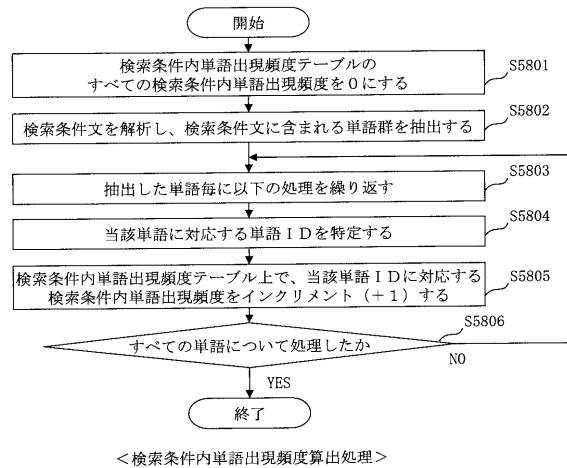
【図55】



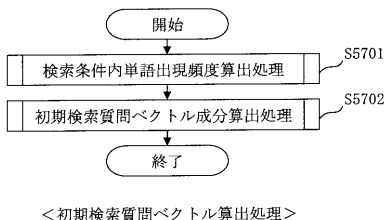
【図56】



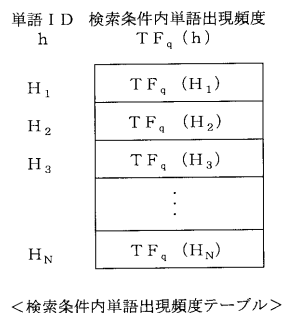
【図58】



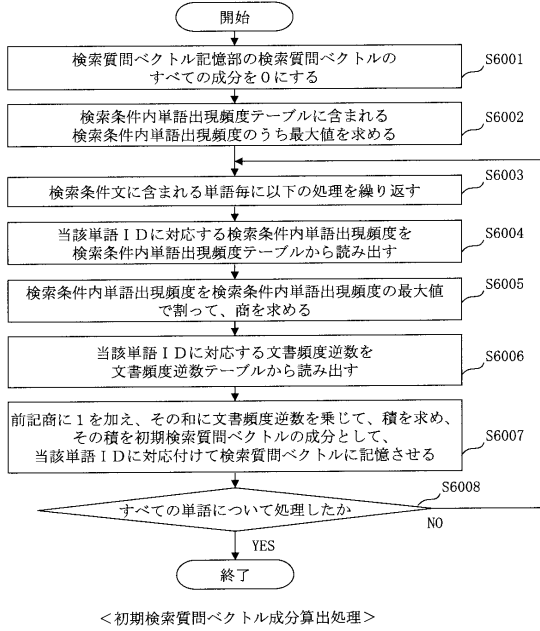
【図57】



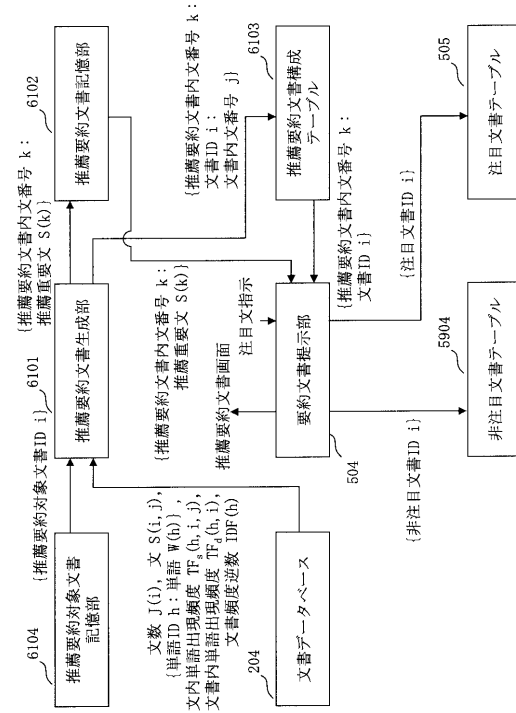
【図59】



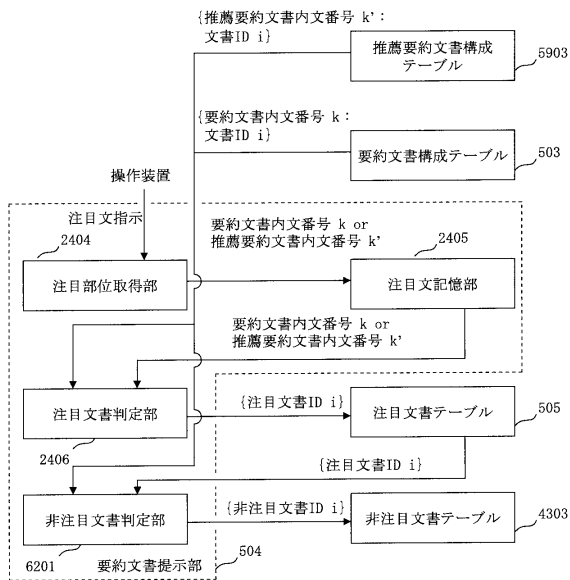
【図60】



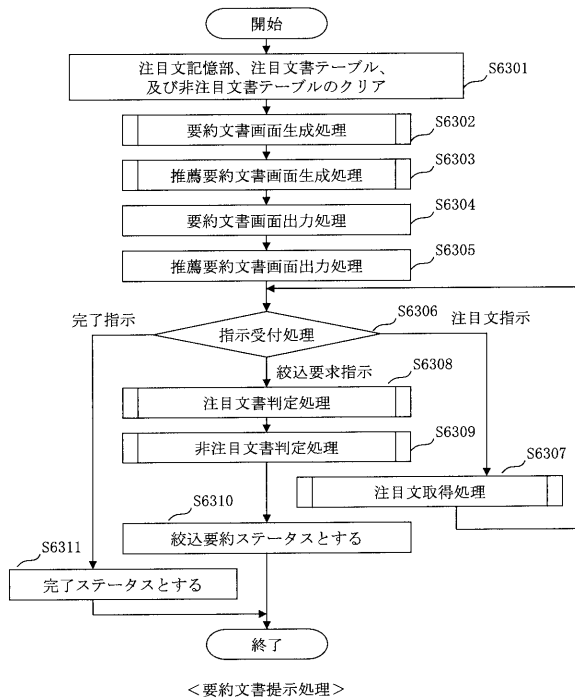
【図61】



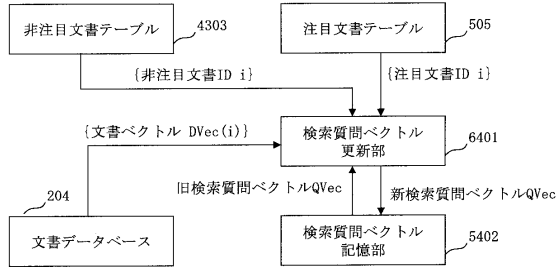
【図62】



【図63】



【図 6 4】



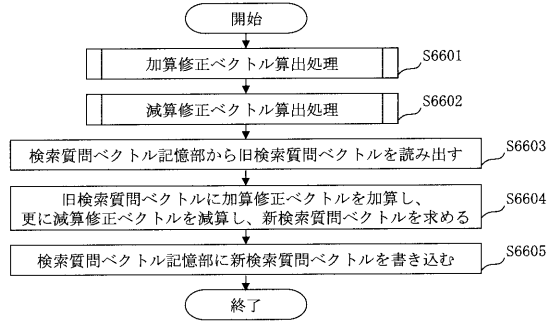
【図 6 5】

文書 ID i

単語 ID h	I ₁	I ₂	I _M
H ₁	W g t _d (H ₁ , I ₁)	W g t _d (H ₁ , I ₂)	⋯ W g t _d (H ₁ , I _M)
H ₂	W g t _d (H ₂ , I ₁)	W g t _d (H ₂ , I ₂)	⋯ W g t _d (H ₂ , I _M)
H ₃	W g t _d (H ₃ , I ₁)	W g t _d (H ₃ , I ₂)	⋯ W g t _d (H ₃ , I _M)
⋮	⋮	⋮	⋮
H _N	W g t _d (H _N , I ₁)	W g t _d (H _N , I ₂)	⋯ W g t _d (H _N , I _M)

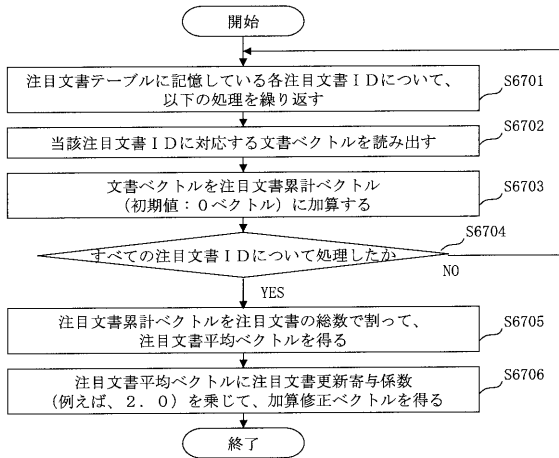
<文書ベクトルテーブル>

【図 6 6】



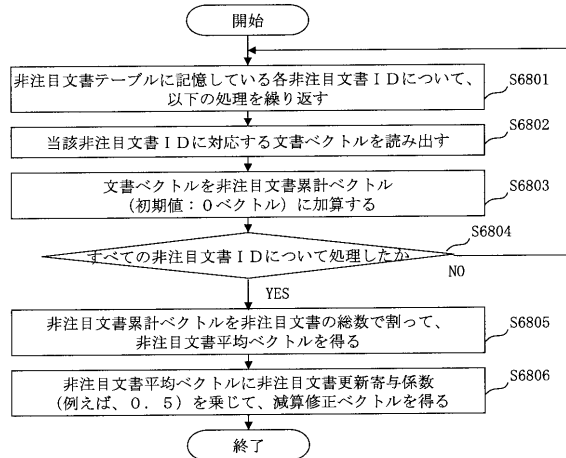
<検索質問ベクトル更新処理>

【図 6 7】



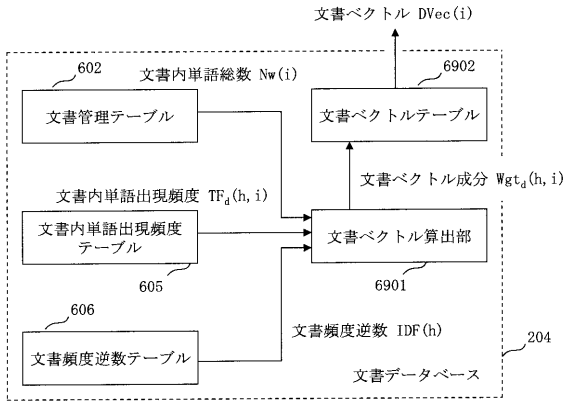
<加算修正ベクトル算出処理>

【図 6 8】

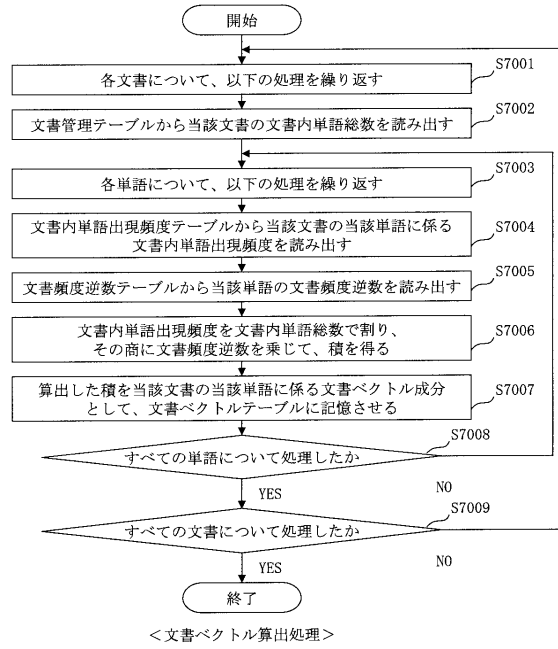


<減算修正ベクトル算出処理>

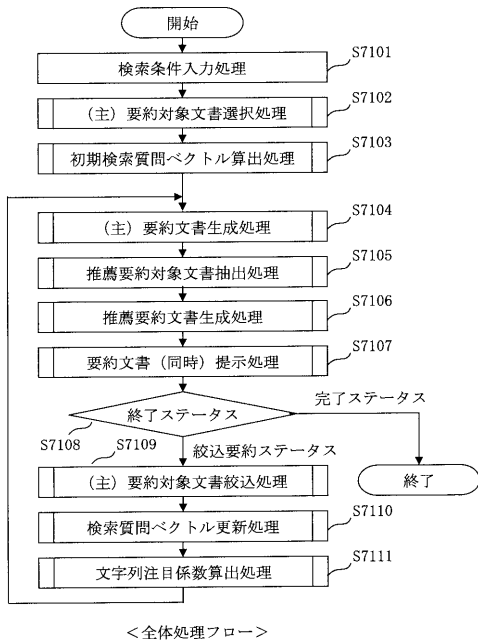
【図 69】



【図 70】

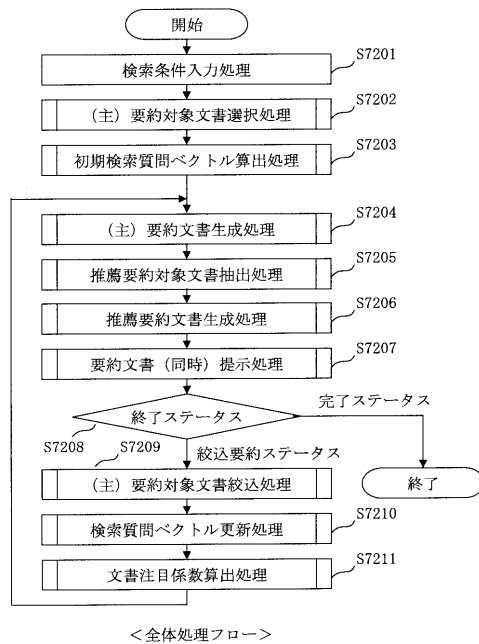


【図 71】



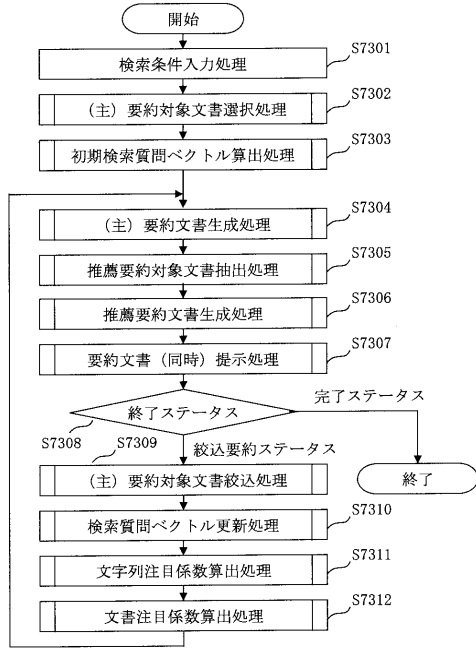
<全体処理フロー>

【図 72】



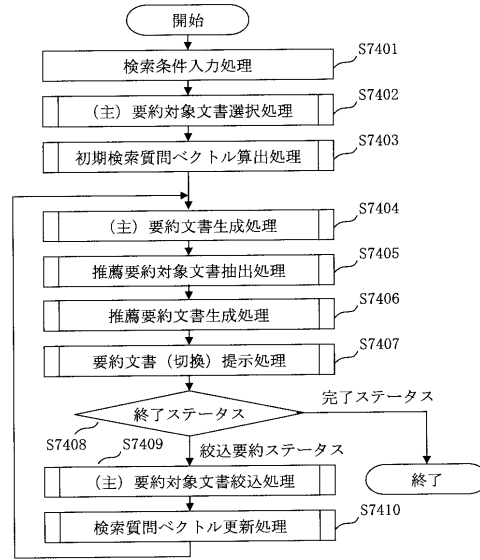
<全体処理フロー>

【図73】



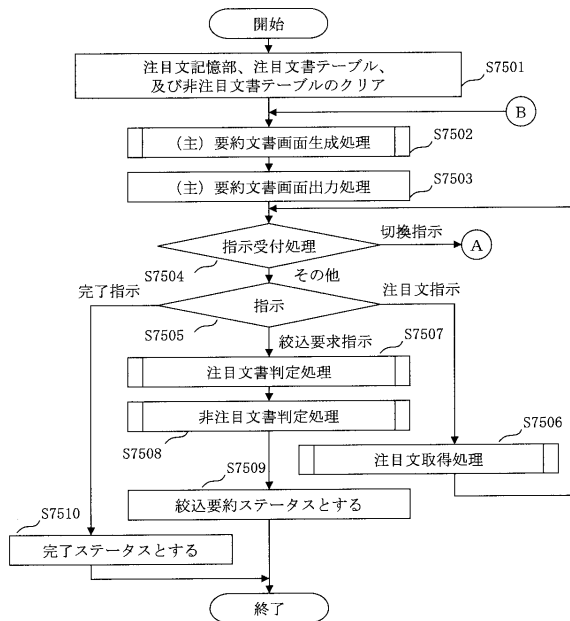
<全体処理フロー>

【図74】



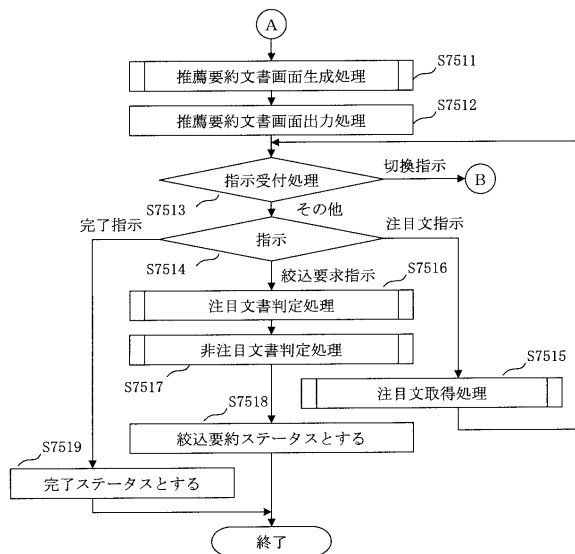
<全体処理フロー>

【図75】



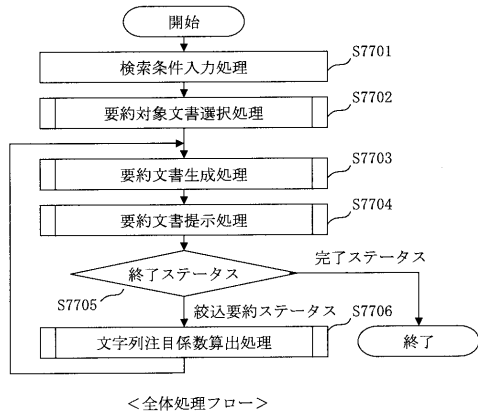
<要約文書（切換）提示処理（1/2）>

【図76】

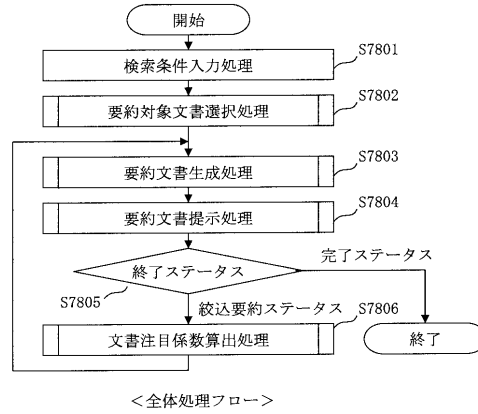


<要約文書（切換）提示処理（2/2）>

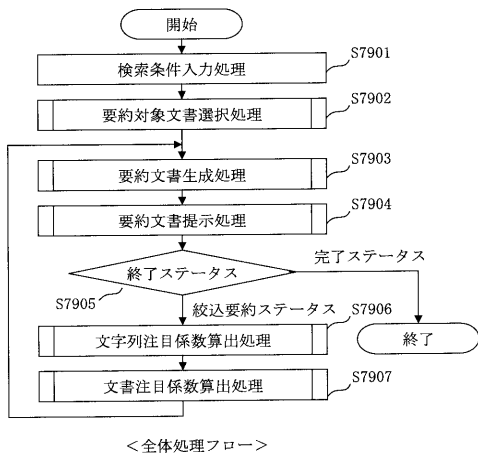
【図77】



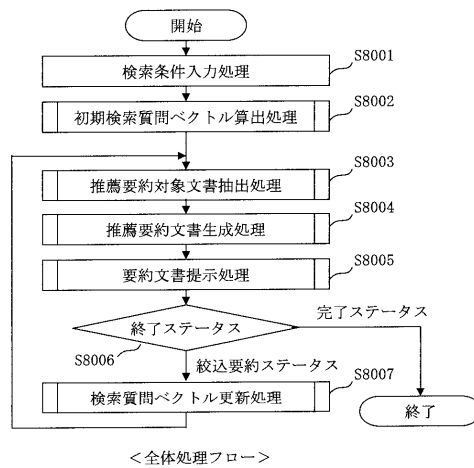
【図78】



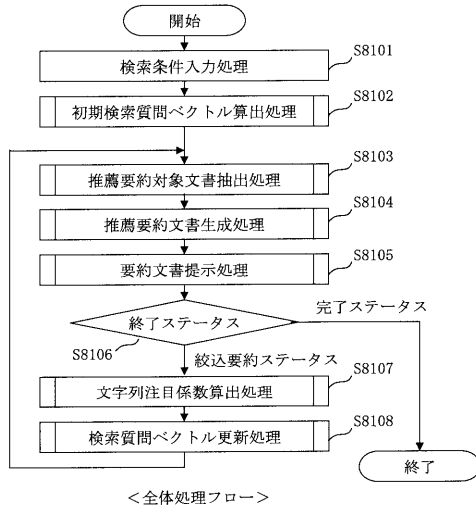
【図79】



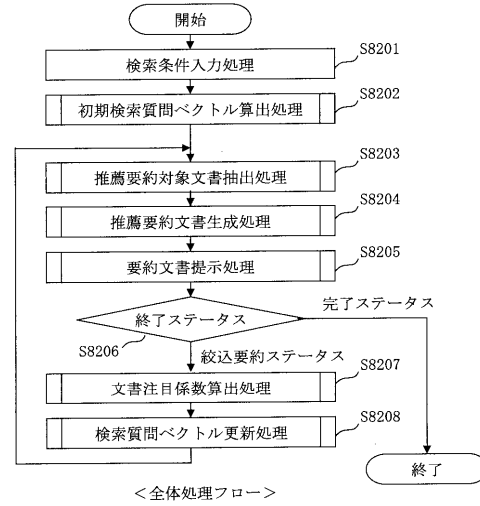
【図80】



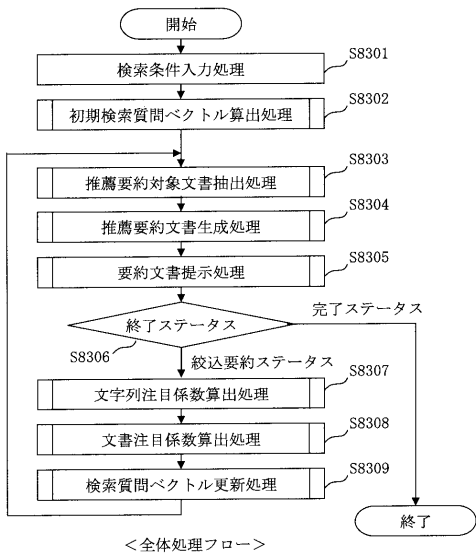
【図 8 1】



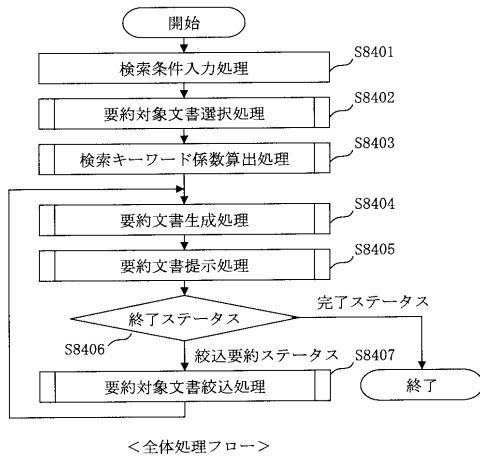
【図 8 2】



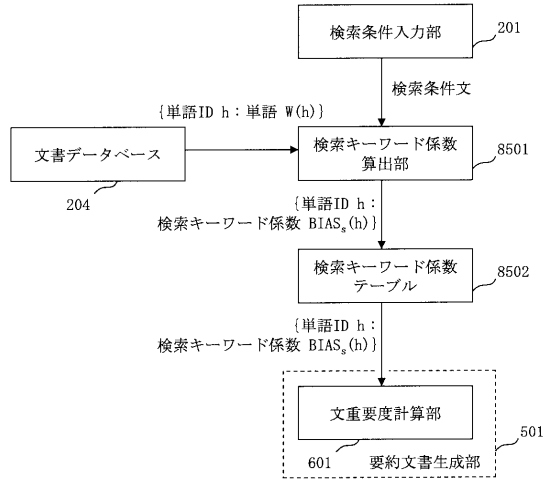
【図 8 3】



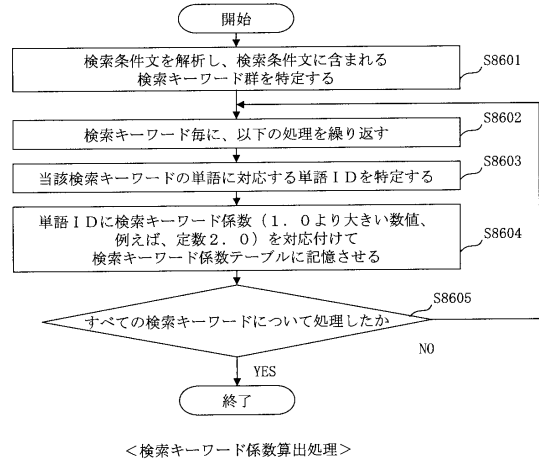
【図 8 4】



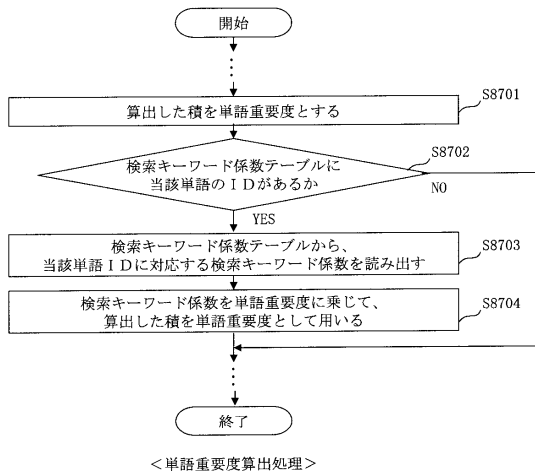
【図 85】



【図 86】



【図 87】



フロントページの続き

- (56)参考文献 特開2006-59082(JP,A)
特開2006-163525(JP,A)
特開2002-041573(JP,A)
特開2003-281159(JP,A)
特開平08-263517(JP,A)
特開2005-141428(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30