

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5127553号

(P5127553)

(45) 発行日 平成25年1月23日(2013.1.23)

(24) 登録日 平成24年11月9日(2012.11.9)

(51) Int.Cl.

F 1

G 0 6 F 17/30 (2006.01)

G 0 6 F 17/30 3 7 0 Z

G 0 6 F 17/30 2 1 0 D

G 0 6 F 17/30 1 7 0 A

請求項の数 7 (全 18 頁)

(21) 出願番号 特願2008-120634 (P2008-120634)
(22) 出願日 平成20年5月2日(2008.5.2)
(65) 公開番号 特開2009-271671 (P2009-271671A)
(43) 公開日 平成21年11月19日(2009.11.19)
審査請求日 平成22年12月20日(2010.12.20)

(73) 特許権者 000006747
株式会社リコー
東京都大田区中馬込1丁目3番6号
(74) 代理人 100070150
弁理士 伊東 忠彦
(72) 発明者 西村 徹郎
東京都大田区中馬込1丁目3番6号 株式
会社リコー内
審査官 伊知地 和之

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法、プログラム及び記録媒体

(57) 【特許請求の範囲】

【請求項1】

文書と該文書の分類を示す分類情報とを関連付けて記憶する記憶手段を有し、前記文書を検索するための検索条件に基づいて前記記憶手段に記憶される文書を検索する情報処理装置であって、

前記検索条件に基づいて検索された検索結果の総数を前記記憶手段から取得する総数取得手段と、

前記検索条件に基づいて検索された所定数の検索結果を前記記憶手段から繰り返し取得する検索結果取得手段と、

前記検索結果取得手段により検索結果が取得される毎に、前記総数取得手段により取得された総数と前記所定数と前記検索結果とに基づいて前記分類情報毎に分類される文書件数の概数を算出する算出手段と、

前記算出手段により算出された概数を前記分類情報毎に表示するための表示データを生成する生成手段と、

ユーザから検索結果を取得する指示を受け付ける受付手段とを備え、

前記検索結果取得手段は、

前記受付手段により指示が受け付けられた場合、残り全ての検索結果を前記記憶手段から取得し、

前記算出手段は、

前記検索結果取得手段により取得された全ての検索結果に基づいて、前記分類情報毎の

10

20

文書件数を求め、

前記生成手段は、

前記算出手段により求められた文書件数を表示するための表示データを生成する情報処理装置。

【請求項 2】

前記算出手段は、前記検索結果取得手段より取得する検索結果に基づいて、前記所定数を累積した累積数に占める前記分類情報毎に分類される文書件数の割合を算出し、該割合と前記総数とを乗算して前記概数を算出する請求項 1 記載の情報処理装置。

【請求項 3】

前記生成手段は、

前記所定数から求められる前記概数の信頼性に係る情報を前記表示データに含める請求項 1 又は 2 記載の情報処理装置。

【請求項 4】

前記生成手段は、

特定の分類情報については、前記算出手段により前記概数が算出されない場合でも前記分類情報を表示するよう表示データを生成する請求項 1 乃至 3 いずれか一項に記載の情報処理装置。

【請求項 5】

文書と該文書の分類を示す分類情報とを関連付けて記憶する記憶手段を有し、前記文書を検索するための検索条件に基づいて前記記憶手段に記憶される文書を検索する情報処理装置における情報処理方法であって、

前記検索条件に基づいて検索された検索結果の総数を前記記憶手段から取得する総数取得ステップと、

前記検索条件に基づいて検索された所定数の検索結果を前記記憶手段から繰り返し取得する検索結果取得ステップと、

前記検索結果取得ステップにより検索結果が取得される毎に、前記総数取得ステップにより取得された総数と前記所定数と前記検索結果とに基づいて前記分類情報毎に分類される文書件数の概数を算出する算出ステップと、

前記算出ステップにより算出された概数を前記分類情報毎に表示するための表示データを生成する生成ステップと、

ユーザから検索結果を取得する指示を受け付ける受付ステップとを有し、

前記検索結果取得ステップは、

前記受付ステップで指示が受け付けられた場合、残り全ての検索結果を前記記憶手段から取得し、

前記算出ステップは、

前記検索結果取得ステップで取得された全ての検索結果に基づいて、前記分類情報毎の文書件数を求め、

前記生成ステップは、

前記算出ステップで求められた文書件数を表示するための表示データを生成する情報処理方法。

【請求項 6】

請求項 5 記載の情報処理方法を、コンピュータに実行させるためのプログラム。

【請求項 7】

請求項 6 記載のプログラムが記録された、コンピュータにより読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法、プログラム及び記録媒体に関し、さらに詳しく言えば、検索された文書の分類情報毎に文書件数を求める情報処理装置、情報処理方法

10

20

30

40

50

、プログラム及び記憶媒体に関する。

【背景技術】

【0002】

従来から、検索機能を有する情報処理装置において、検索キーワードをもとに、データベースから文書を検索し、検索結果を表示する場合、文書のある視点から捉えた分類情報毎に文書件数を表示する手法が提案されている。

【0003】

例えば、特開平07-319905（特許文献1）には、検索キーによって検索された検索結果を表示する情報検索装置に関し、検索キーに一致するデータを複数の観点から分類して統合した分類結果をもとに、複数の観点からの分類結果を同時に識別できるように表示し、検索結果の絞込みを有効に支援する技術が開示されている。

10

【特許文献1】特開平07-319905号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかし、特許文献1の技術では、検索結果を複数の観点から分類するとき、検索結果全てに対して分類を行なうため、検索結果が多ければ多いほど検索結果の分類に時間がかかってしまうため、分類情報の表示に時間がかかってしまうという問題点があった。

【0005】

本発明は、上記問題に鑑みて、検索結果を分類する場合、検索結果の分類情報に関する表示処理を高速に行うことができる情報処理装置、情報処理方法、プログラム及び記録媒体を提供することを目的とする。

20

【課題を解決するための手段】

【0006】

本発明の一局面の情報処理装置は、文書と該文書の分類を示す分類情報とを関連付けて記憶する記憶手段を有し、前記文書を検索するための検索条件に基づいて前記記憶手段に記憶される文書を検索する情報処理装置であって、前記検索条件に基づいて検索された検索結果の総数を前記記憶手段から取得する総数取得手段と、前記検索条件に基づいて検索された所定数の検索結果を前記記憶手段から繰り返し取得する検索結果取得手段と、前記検索結果取得手段により検索結果が取得される毎に、前記総数取得手段により取得された総数と前記所定数と前記検索結果とに基づいて前記分類情報毎に分類される文書件数の概数を算出する算出手段と、前記算出手段により算出された概数を前記分類情報毎に表示するための表示データを生成する生成手段と、ユーザから検索結果を取得する指示を受け付ける受付手段とを備え、前記検索結果取得手段は、前記受付手段により指示が受け付けられた場合、残り全ての検索結果を前記記憶手段から取得し、前記算出手段は、前記検索結果取得手段により取得された全ての検索結果に基づいて、前記分類情報毎の文書件数を求め、前記生成手段は、前記算出手段により求められた文書件数を表示するための表示データを生成する。

30

【0007】

また、本発明の他の局面の情報処理方法は、文書と該文書の分類を示す分類情報とを関連付けて記憶する記憶手段を有し、前記文書を検索するための検索条件に基づいて前記記憶手段に記憶される文書を検索する情報処理装置における情報処理方法であって、前記検索条件に基づいて検索された検索結果の総数を前記記憶手段から取得する総数取得ステップと、前記検索条件に基づいて検索された所定数の検索結果を前記記憶手段から繰り返し取得する検索結果取得ステップと、前記検索結果取得ステップにより検索結果が取得される毎に、前記総数取得ステップにより取得された総数と前記所定数と前記検索結果とに基づいて前記分類情報毎に分類される文書件数の概数を算出する算出ステップと、前記算出ステップにより算出された概数を前記分類情報毎に表示するための表示データを生成する生成ステップと、ユーザから検索結果を取得する指示を受け付ける受付ステップとを有し、前記検索結果取得ステップは、前記受付ステップで指示が受け付けられた場合、残り全

40

50

ての検索結果を前記記憶手段から取得し、前記算出ステップは、前記検索結果取得ステップで取得された全ての検索結果に基づいて、前記分類情報毎の文書件数を求め、前記生成ステップは、前記算出ステップで求められた文書件数を表示するための表示データを生成する。

【発明の効果】

【0008】

本発明によれば、検索結果を分類する場合、検索結果の分類情報に関する表示処理を高速に行うことができる。

【発明を実施するための最良の形態】

【0009】

以下、本発明の実施例を図面に基づいて説明する。

【実施例1】

【0010】

＜システムの概略ハードウェア構成について＞

まず、本発明に係るシステムの概略ハードウェア構成について説明する。図1は、本発明に係るシステムの概略ハードウェア構成の一例を示す図である。

【0011】

図1に示すように、本発明に係るシステムは、ネットワークを介して情報処理装置100とクライアント105、106とが接続され、情報処理装置100は、CPU (Central Processing Unit) 101、HDD (Hard Disk Drive) 102、RAM (Random Access Memory) 103、ネットワークI/F 104を含んで構成される。

【0012】

CPU 101は、コンピュータの中で、各装置の制御やデータの演算、加工を行うCPUである。CPU 101は、RAM 103に記憶されたプログラムを実行する演算装置で、入力装置や記憶装置からデータを受け取り、演算、加工した上で、出力装置や記憶装置に出力する。

【0013】

HDD 102は、アプリケーションソフトウェアなどに関連するデータを記憶する記憶装置である。また、HDD 102には、文書（テキストデータと同義）データが記憶されており、文書が検索されるたびに文書データが読み出されたり、新規に文書データが書き込まれたりする。文書データとは、文書IDやタイトル、テキストデータ、分類情報などの総称である。

【0014】

RAM 103は、CPU 101が実行する基本ソフトウェアであるOS（オペレーティングシステム）やアプリケーションソフトウェアなどのプログラムやデータを記憶又は一時保存する記憶装置である。

【0015】

ネットワークI/F 104は、有線及び／又は無線回線などのデータ伝送路により構築されたLAN (Local Area Network)、WAN (Wide Area Network)などのネットワークを介して接続された通信機能を有するクライアント105、106と当該情報処理装置のインタフェースである。

【0016】

クライアント105、106は、情報処理装置100にネットワークを介して接続された情報処理端末であり、例えば、PC (Personal Computer) やPDA (Personal Digital Assistant)、携帯端末などである。また、図1に示す例においてクライアント2つが、情報処理装置に接続されているが、クライアントは1つ又は複数情報処理装置に接続されるようにしてもよい。

【0017】

以上のように構成されたシステムにおいて、情報処理装置100は、クライアント105、106から検索文字列などが入力され、取得した検索文字列などに基づく検索条件に

10

20

30

40

50

より検索を行い、検索結果をクライアントに表示させる。なお、情報処理装置 100 は、検索文字列の入力から検索結果の表示まで一連の検索処理全てを行なうように構成してもよい。

【0018】

＜実施例 1 に係る情報処理装置とクライアントの主要機能構成について＞

実施例 1 に係る情報処理装置とクライアントの主要機能構成について説明する。図 2 は、実施例 1 に係る情報処理装置とクライアントの主要機能構成を示すブロック図である。

【0019】

図 2 に示すように、実施例 1 の情報処理装置 100 は、検索条件生成手段 203、検索結果取得手段 204、分類情報取得手段 205、検索結果総数取得手段 206、記憶手段 207、分類結果生成手段 208 を含んで構成される。クライアント 105（クライアント 106 も同様）は、検索文字列入力手段 201、分類条件取得手段 202、検索結果表示手段 209、分類結果表示手段 210 を含んで構成される。

10

【0020】

まず、情報処理装置 100 の構成について説明する。検索条件生成手段 203 は、クライアント 105 側から入力される検索文字列を検索条件とし、検索結果取得手段 204、分類情報取得手段 205、検索結果総数取得手段 206 に検索条件を出力する。

【0021】

また、検索条件生成手段 203 は、後述するファセットや分類情報を、分類条件指定手段 202 から入力された場合は、取得した検索文字列とファセットや分類情報とに基づいて検索条件を生成し、生成した検索条件を検索結果取得手段 204、分類情報取得手段 205、検索結果総数取得手段 206 に出力する。

20

【0022】

ここで、検索文字列とは、単語やキーワード、文章などであり、数字や記号なども含めた文字列であるとする。なお、文章が入力された場合は、文章を形態素解析するなどして単語に分割し、分割された単語に基づいて検索が行なわれる。

【0023】

検索結果取得手段 204 は、検索条件生成手段 203 より取得した検索条件に基づいて記憶手段 207 にアクセスして問合せを行い、検索結果を取得する。このとき、記憶手段 207 は、検索対象の文書に対し、全文索引を付与して記憶しているとし、検索条件に含まれる検索文字列を含むテキストデータなどが検索結果として検索条件取得手段 204 に取得される。

30

【0024】

また、検索結果取得手段 204 は、検索条件に分類情報が含まれている場合は、B-tree 索引を用いて検索を行なう。ただし、分類情報の異なり数が少ない場合は、B-tree 索引を利用しない。なお、記憶手段 207 に記憶される文書データについては図 3 に示すが、詳細は後述する。

【0025】

また、検索結果取得手段 204 は、検索結果を所定数 n ずつ記憶手段 207 から繰り返し取得するとし、検索結果を取得する度に検索結果をクライアント側の検索結果表示手段 209 に送信する。

40

【0026】

なお、所定数 n は、ユーザにより適宜設定されてもよいし、予め設定された値を用いるようにしてもよい。

【0027】

分類情報取得手段 205 は、検索条件生成手段 203 より取得した検索条件に基づいて記憶手段 207 にアクセスして問合せを行い、分類情報を取得する。ここで、分類情報は、メタデータとして文書データに関連付けられて記憶手段 207 に記憶されているとしてもよい。

【0028】

50

また、分類情報とは、文書の分類を示し、ファセット毎に与えられる。ファセットとは、文書の観点（視点）を示し、例えば、文書の種別、出版年、出版社などの情報である。分類情報は、例えば、ファセットが「文書の種別」である場合の文書情報は、書籍、論文、新聞などであり、ファセットが「出版年」である場合の文書情報は、2006年、2005年などの年度情報である。

【0029】

また、分類情報取得手段205は、検索条件に含まれる検索文字列をテキストデータを含む文書データのメタデータを、検索結果取得手段204同様、所定数nずつ記憶手段207から繰り返し取得する。また、分類情報の検索においては、分類情報がユーザにより指定された場合は、B-tree索引を用いて検索を行う。ただし、分類情報の異なり数が少ない場合はB-tree索引を利用しない。

10

【0030】

次に、分類情報取得手段205は、繰り返し取得したメタデータから分類情報を取得し、分類情報の文書件数をカウントし、メタデータを取得する度に、分類情報毎にカウント値を分類結果生成手段208に出力する。

【0031】

ここで、分類情報取得手段205の処理について図4と図5とを用いて説明する。図4は、所定数nを5としたとき、記憶手段207から取得したメタデータに基づく分類情報の一例を示す図である。

【0032】

20

図4に示すように、検索条件に含まれる検索文字列を含むテキストデータが5つ（文書IDが1、2、5、10、16）記憶手段207から読み出されている。また、文書IDに関連付けてファセット毎（種別、出版年、出版社など）に分類情報が記憶手段207から読み出されている。

【0033】

例えば、図4に示す例では、文書ID「1」のファセット「種別」に対する分類情報は「書籍」であり、ファセット「出版年」に対する分類情報は「2006」であり、ファセット「出版社」に対する分類情報は「A社」であることを示している。

【0034】

図5は、ファセットに着目したときの各分類情報の件数の一例を示す図である。図5に示すように、図4に示す分類情報をファセット毎に着目して各分類情報の件数をカウントする。

30

【0035】

例えば、図5（A）は、図4に示す5つのデータを、ファセット「種別」に着目したときの各分類情報（書籍、論文など）の件数をカウントした結果を示す例である。図5（B）は、ファセット「出版年」に着目したときの各分類情報（2006、2005など）の件数をカウントした結果を示す例である。図5（C）は、ファセット「出版社」に着目したときの各分類情報（A社、B社）の件数をカウントした結果を示す例である。

【0036】

図2に戻り、検索結果総数取得手段206は、検索条件生成手段203より取得した検索条件に基づいて、検索条件に含まれる検索文字列を含むテキストデータが、記憶手段207に全部で何件記憶されているかをカウントし、カウント値を総数として分類結果生成手段208に出力する。

40

【0037】

記憶手段207は、検索対象の文書を分類情報などに関連付けて記憶する。テキストデータには、前述したように全文索引が付与され、さらに、ここでは、登録日、分類情報などのメタデータも付与されて記憶されているとする。これより、高速に検索ができるようになっている。

【0038】

ここで、図3は記憶手段に記憶される文書データの一例を示す図である。図3に示すよ

50

うに、文書データは、文書ID、タイトル、テキストデータ、種別、出版年、出版社などの情報が関連付けられて記憶されている。ここで、文書の観点を示すファセットは、種別、出版年、出版社などであり、文書の分類を示す分類情報は、書籍、論文、2005、2006、A社、B社などである。

【0039】

また、記憶手段207は、分類情報に対して、例えばB-tree索引を付して記憶してもよい。このとき、全ての分類情報に対して索引を付すのではなく、パラメータの異なり数が多いものに対してのみ索引を付す。パラメータの異なり数が少ないものに索引を付しても高速に検索ができない場合があるからである。

【0040】

図2に戻り、分類結果生成手段208は、分類情報取得手段205より取得した分類情報毎のカウント値と、検索結果総数取得手段206より取得した検索結果の総数のカウント値と、所定数 n とに基づいて、分類情報毎に分類される文書件数の概数を算出する。また、分類結果生成手段208は、算出した概数を分類情報毎に表示するため表示データを生成する。

【0041】

ここで、分類結果生成手段208の詳細な機能について図6を用いて説明する。図6は、分類結果生成手段208の詳細な機能構成を示すブロック図である。分類結果生成手段208は、概数算出手段601と表示データ生成手段602とを含んで構成される。

【0042】

概数算出手段601は、まず、分類情報取得手段205より取得した分類情報毎のカウント値と、検索結果総数取得手段206より取得した検索結果の総数のカウント値と、所定数 n とを取得する。

【0043】

所定数 n に関しては、分類情報毎のカウント値の総和が所定数 n となるので、分類情報毎のカウント値を累積加算して取得してもよいし、分類情報取得手段205から取得するようにしてもよい。

【0044】

次に、概数算出手段601は、取得した所定数 n に占める分類情報毎のカウント値の割合を算出する。この割合は、分類情報毎のカウント値を所定数 n で除算することにより求められる。

【0045】

次に、概数算出手段601は、求めた割合に総数のカウント値を乗算した結果を、分類情報毎の概数として算出し、算出した分類情報毎の概数を表示データ生成手段602に出力する。なお、前述した算出の手順は問わず、結果的に、同じ値が算出されるのであれば、分類情報毎のカウント値に総数を乗算した後、所定数 n で除算するようにしてもよい。

【0046】

また、概数算出手段601は、分類情報取得手段205から分類情報毎のカウント値を取得する度に、所定数 n と分類情報毎のカウント値とを累積加算して、前述したように概数を算出し、概数を算出する度に表示データ生成手段602に出力する。

【0047】

表示データ生成手段602は、概数算出手段301より取得した分類情報毎の概数を、分類情報に関連付けて表示画面に表示されるよう表示データを生成し、生成した表示データをクライアント側の分類結果表示手段210に送信する。

【0048】

また、表示データ生成手段602は、概数を取得する度に、前述した処理を行い、分類結果表示手段210に表示データを出力する。なお、表示データ生成手段602は、表示データを生成する際には、概数が整数で表示されるよう概数を整数に丸め込んで表示データを生成するようにすればよい。

【0049】

10

20

30

40

50

図2に戻り、クライアント105の構成について説明する。検索文字列入力手段201は、キーボードやタッチパネルなどの入力装置を通して検索を行なうための文字列を入力する。また、検索文字列入力手段201は、入力された検索文字列を情報処理装置側の検索条件生成手段203に送信する。

【0050】

分類条件指定手段202は、分類結果表示手段210により表示された分類情報やファセットがユーザにより選択された場合（ユーザが表示画面上の分類情報などをクリック、押下した場合）、選択された分類情報等を情報処理装置側の検索条件生成手段203に送信する。

【0051】

検索結果表示手段209は、情報処理装置側の検索結果取得手段204より検索結果を取得すると、取得した検索結果を表示画面に表示する。

【0052】

分類結果表示手段210は、情報処理装置側の分類結果生成手段208より表示データを取得すると、取得した表示データを表示画面に表示する。なお、分類情報はユーザによりクリック、又は押下可能なように表示画面に表示される。

【0053】

図7は、実施例1における表示画面の一例を示す図である。図7に示すように、表示画面は、検索文字列を入力する表示領域701と、検索結果表示手段209による表示領域702と、分類結果表示手段210による表示領域703とを含んで構成される。

【0054】

表示領域701は、検索文字列を入力する表示領域である。検索文字列は、キーボードやマウス、タッチパネルなどの入力装置により入力される。

【0055】

表示領域702は、検索結果表示手段209によって表示される領域である。図7に示す例では、文書のタイトルとテキストデータの一部が表示されている。また、表示領域702には、他にも、文書のアイコンを表示したり、文書が書籍などの場合には、表紙を表示したりしてもよい。

【0056】

表示領域703は、分類結果表示手段210によって表示される領域である。図7に示す例では、ファセット毎に分類情報が表示され、分類情報には文書件数が表示されている。なお、分類情報については、ユーザによりクリック、又は押下できるように構成されており、クリック、又は押下された分類情報は、検索条件生成手段203に送信されるように構成されている。

【0057】

＜実施例1における検索結果表示処理について＞

図8は、実施例1における検索結果表示処理のフローチャートである。ステップ801では、検索結果総数取得手段206が、検索結果の総数であるtotalを取得し、分類結果生成手段208にtotalを出力する。

【0058】

ステップ801に続いてステップ802に進み、検索結果取得手段204が、検索の開始位置であるstartからstart+n-1件の検索結果を、全文索引などを用いて取得し、取得した検索結果を検索結果表示手段209に出力する。このあと、検索結果は、検索結果表示手段209により表示画面に表示される。

【0059】

このとき、検索結果取得手段209は、ユーザにより入力された文字列で最初に検索を行なうときは、全文索引を用いて検索を行う。次に、検索結果取得手段209は、ユーザにより分類情報などが指定され、検索結果を絞り込むときは、B-tree索引を用いて検索結果を絞り込む。ただし、分類情報の異なり数が少ない場合は、B-tree索引を利用しない。

10

20

30

40

50

【0060】

なお、startの初期値は、検索条件に一致する文書データの中で最初の文書からということで1が設定されているとする。また、件数nは、ユーザにより取得したい件数を設定されるようにしてもよいし、情報処理装置で適宜設定されるようにしてもよい。

【0061】

ステップ802に続いてステップ803に進み、分類情報取得手段205が、全文索引などを用いて検索を行い、startからstart+n-1件の分類情報を取得する。このとき、ユーザにより分類情報が指定された場合は、B-tree索引を用いて分類情報を取得してもよい。なお、前述したように、分類情報の異なり数が少ない場合は、B-tree索引を利用しない。

10

【0062】

なお、ステップ802、803において、取得する件数がn件に満たない場合がある。この場合には、n件未満の取得可能な件数を取得することにする。なお、取得する件数がn件に満たない条件としては、検索条件に一致する文書が、n件未満の場合又はn件の倍数ではない場合が考えられる。

【0063】

ステップ803に続いてステップ804に進み、分類情報取得手段205が、取得した分類情報のうち、分類情報vごとに同じ分類情報が何件あるかをカウントすることで文書件数f(v)を取得する。

【0064】

20

ステップ804に続いてステップ805に進み、分類情報取得手段205が、facet(v)にf(v)を加算し、facet(v)を概数算出手段601に出力する。なお、facet(v)には、初期値として0が設定されている。これより、分類情報毎の文書件数を累積加算していくことになる。

【0065】

ステップ805に続いてステップ806に進み、概数算出手段205が、start+n ≥ totalか否かを判定する。

【0066】

ステップ806の判定結果がNOである場合、ステップ807に進み、概数算出手段807が、

$$total \times facet(v) / (start + n - 1) \quad \cdots \text{式(1)}$$
 を計算して分類情報毎の文書件数の概数を算出し、算出した概数を表示データ生成手段602に出力する。次に表示データ生成手段602は、取得した概数に基づいて表示データを生成し、生成した表示データを分類結果表示手段210に出力する。次に、分類結果表示手段210が、取得した表示データを表示画面に表示する。

30

【0067】

ステップ807に続いてステップ808に進み、検索結果取得手段204、分類情報取得手段205は、

start = start + n $\cdots$ 式(2)

を計算し、startの開始位置を更新し、ステップ802に戻る。

40

【0068】

ステップ806の判定結果がYESである場合、ステップ809に進み、概数算出手段601が、facet(v)は、概数ではなく正確な件数であると判断して、facet(v)を表示データ生成手段602に出力し、前述したようにfacet(v)が表示画面に表示される。

【0069】

なお、ステップ801、ステップ802、ステップ803の各処理については、必ずしもこの順序で行う必要はなく、並列に処理してもよいし、順序は特に関係ない。

【0070】

以上、実施例1に係る情報処理装置によれば、検索結果を分類する場合、分類情報の文

50

書件数の概数を算出することで、検索結果の分類情報に関する表示処理を高速に行うことができる。

【0071】

また、図8に示す処理において、ステップ806の判定結果がYESとなるまで、表示画面に分類情報の文書件数に対して更新中を示す情報を表示するようにしてもよい。例えば、ステップ806の判定結果がNOである場合は、「更新中」という文字を表示画面に表示させ、ステップ806の判定結果がYESとなった場合に「更新終了」という文字を表示画面に表示させるようにしてもよい。これより、ユーザは、分類情報の文書件数がまだ更新中であることを視覚的に理解することができる。

【0072】

また、ステップ807の後に、所定時間待機する処理を付加してもよいし、ユーザによりn件取得する指示がなされたかどうかの判定を行なって、指示がなされたときにステップ808の指示を行なうようにしてもよい。これより、次のn件の情報を必要とするときだけ、次のn件の情報を取得することができ、ユーザにとって無駄な処理を省くことができる。

【0073】

なお、実施例1では、情報処理装置100とクライアント105、106は区別した構成にしたが、これらを一体として構成し、検索文字列の入力から検索結果の表示まで全て1つの装置によって行うようにしてもよい。

【0074】

[変形例1]

実施例1に係る情報処理装置の変形例1について説明する。変形例1に係る情報処理装置は、ユーザにより、残り全ての検索結果を取得する指示を受けた場合に、残り全ての検索結果を取得して、分類情報の正確な文書件数を表示させる。

【0075】

＜変形例1に係る情報処理装置とクライアントの主要機能構成について＞

図9は、変形例1に係る情報処理装置とクライアントの主要機能構成を示すブロック図である。なお、図9に示す構成要素において、図2に示す構成要素と同様の機能のものは、図2と同じ符号を付し、その説明を省略する。

【0076】

図9に示すように、クライアント905は、図2に示すクライアント104と比較して新たに検索結果取得指示手段901が追加されている。検索結果取得指示手段901は、ユーザにより残り全ての検索結果を取得するためにクリック、又は押下されるボタンなどである。

【0077】

図10は、変形例1における表示画面の一例を示す図である。図10に示す取得ボタン1001は、検索結果取得手段901であり、この取得ボタン1001がクリック、又は押下されると、指示受付手段902にその旨通知する。また、図10に示すように、変形例1ではファセット1002である種別、出版年、出版社なども、これらの正確な文書件数を取得するためにクリック、又は押下可能となっている。

【0078】

図9に戻り、情報処理装置900は、図2に示す情報処理装置100と比較して新たに指示受付手段902が追加されている。指示受付手段902は、検索結果取得指示手段901より通知を受けると、検索結果取得手段903、分類情報取得手段904に残り全ての検索結果、分類情報をそれぞれ取得するよう指示する。

【0079】

検索結果取得手段903、分類情報取得手段904は、検索結果総数取得手段206よりtotalを取得し、nにtotal-startを代入する。これより、次に検索結果、分類情報を取得するときは、残り全ての情報を取得することができる。残り全ての情報を取得した後の処理は、件数が違うのみで図2と同様の処理を行なう。

【0080】

なお、指示受付手段902は、検索結果取得手段903、分類情報取得手段904に指示を出すのではなく、検索結果総数取得手段206に検索結果取得手段903、分類情報取得手段904に総数を出力するよう指示を出してもよい。このとき、検索結果総数取得手段206は、指示受付手段902から指示を受けた場合、総数を検索結果取得手段903、分類情報取得手段904に出力する。

【0081】

ここで、図10に示すように、ファセット1002がクリック、又は押下された場合には、クリック、又は押下されたファセットのみ、残り全ての検索結果に基づく正確な分類情報毎の文書件数が表示されるようにしてもよい。

10

【0082】

例えば、ファセット「種別」がクリック、又は押下された場合には、分類結果の表示領域703には、「種別」の分類情報（書籍、論文、新聞、雑誌）のみを表示し、表示された分類情報毎に、全ての検索結果に基づく正確な文書件数を表示するようにしてもよい。

【0083】

＜変形例1における検索結果表示処理の一部について＞

変形例1における検索結果表示処理の一部について説明する。図11は、変形例1における検索結果表示処理の一部のフローチャートである。具体的には、図11は、ユーザから検索結果取得指示があった場合のフローチャートである。なお、ユーザから検索結果取得指示がなかった場合は図8に示す処理と同様であるため省略する。

20

【0084】

ステップ1101では、指示受付手段902が、検索結果取得指示手段901から、検索結果を取得する旨の通知が来たか否かを判定する。

【0085】

ステップ1101の判定結果がNOである場合は、ステップ1101に戻る。ステップ1101の判定結果がYESである場合は、ステップ1102に進み、分類情報取得手段904が残り全ての分類情報vを取得し、検索結果取得手段903が、残り全ての検索結果を取得する。

【0086】

ステップ1102に続いてステップ1103に進み、分類情報取得手段904は、取得した残り全てのデータのうち、同じ分類情報vが何件あるかをカウントすることで文書件数f(v)を取得する。

30

【0087】

ステップ1103に続いてステップ805に進み、後は図8に示す処理と同様の処理を行なう。

【0088】

以上、変形例1に係る情報処理装置によれば、ユーザから残り全ての検索結果を取得する指示を受けた場合に、残り全ての検索結果、分類情報を取得することにより、ユーザに対し、分類情報の文書件数の中間値を省いて、正確な文書件数を表示画面に表示させることができる。変形例1は、ユーザが分類情報の文書件数の概数を必要とせず、正確な文書件数を必要とするときに有効である。

40

【0089】

[変形例2]

実施例1に係る情報処理装置の変形例2について説明する。変形例2では、文書件数の概数の信頼性に係る情報を表示画面に表示する。

【0090】

図12は、変形例2における表示画面の一例を示す図である。図12に示すように、表示領域1201が文書件数の概数の信頼性に係る情報を表示する領域である。図12に示す例では、totalが12345件であり、現在2000件に基づいて文書件数の概数を算出していることを示している。

50

【0091】

これは、表示データ生成手段602が、所定数nの累積数とtotalに基づいて、表示画面における所定の位置に、文書件数の概数の信頼性に係る情報を表示させるよう表示データを生成すればよい。

【0092】

なお、表示領域1201に表示する情報としては、概数を算出した実際の件数（n件の累積件数）やtotalに対するパーセンテージ（%）などでもよい。

【0093】

以上、変形例2に係る情報処理装置によれば、分類結果の表示領域に表示されている文書件数の概数が、どれくらい信頼できるものなのかを視覚的に理解することができる。これより、ユーザは、信頼性に係る情報を見ることにより、十分な件数に基づく文書件数の概数だと判断すれば、分類情報をクリックするなどしてさらに条件を絞ればよいし、不十分な件数に基づく文書件数の概数だと判断すれば、さらに、検索結果を取得するようにすればよい。

10

【0094】

〔変形例3〕

実施例1に係る情報処理装置の変形例3について説明する。変形例3では、特定の分類情報について、文書件数の概数が算出されなくても、この特定の分類情報は表示画面に表示されるようにする。つまり、変形例3では、表示画面に対して常に表示される固定の分類情報を設定する。

20

【0095】

図13は、変形例3における表示画面の一例を示す図である。図13に示すように、ファセット「種別」の分類情報1301は、常に表示領域703に表示されるように設定されている。よって、図13に示すように、文書件数の概数を表示している段階であっても、分類情報「雑誌」のように0件を表示させることができる。

【0096】

これは、表示データ生成手段602が、予め特定の分類情報を記憶しておき、この分類情報に対しては、概数算出手段601より概数を取得した場合、概数を表示するようにし、概数を取得しなかった場合、「0」を表示するようにする。なお、特定の分類情報については、情報処理装置の管理者などが予め設定しておけばよい。

30

【0097】

また、変形例3において、概数を表示する段階で、文書件数が0件の分類情報もユーザによりクリック、又は押下可能なように構成する。これは、現時点では文書件数は0件かもしれないが、残りの検索結果によっては、この分類情報に分類される文書が存在するかも知れないからである。これより、概数を表示する段階で文書件数が0件の分類情報をユーザがクリック、又は押下した場合に、この分類情報は本当に文書件数が0件なのかどうかを、ユーザはいち早く知ることができる。

【0098】

なお、全ての検索結果を分類した結果も文書件数が0件である場合は、この分類情報は、0件であることがわかった時点でユーザによりクリック、又は押下不可能なようにすればよい。これより、ユーザによる無意味な検索条件の絞込み、誤操作を防ぐことができる。

40

【0099】

以上、変形例3に係る情報処理装置によれば、特定の分類情報に対しては、概数が算出されていなくても表示画面に表示されるようにしたことにより、特に重要なファセットに対していかなる分類情報があるかをユーザに表示することができる。

【0100】

〔変形例4〕

実施例1に係る情報処理装置の変形例4について説明する。変形例4では、検索結果取得手段204に分類情報取得手段205を一体化して構成する。

50

【0101】

変形例4では、検索結果取得手段204は、記憶手段207からn件ずつ検索結果として文書データを繰り返し読み出して取得する。検索結果取得手段204は、取得したn件の検索結果から、表示領域702に必要なデータ（タイトル、テキストデータの一部）のみを抽出し、抽出したデータを検索結果表示手段209に出力する。

【0102】

次に、検索結果取得手段204は、取得したn件の検索結果から、分類情報毎に、同じ分類情報が何件あるかをカウントする。カウントした値は、分類情報毎に分類結果生成手段208に出力する。上記処理を、記憶手段207から検索結果をn件取得する度に行なう。

10

【0103】

以上、変形例4に係る情報処理装置は、検索結果取得手段と分類情報取得手段とを一体化した構成であっても、検索結果を分類する場合、分類情報の文書件数の概数を算出することで、検索結果の分類情報に関する表示処理を高速に行うことができる。

【0104】

以上、本発明の実施例について詳述したが、本発明に係る特定の実施例に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、上記変形例以外にも種々の変形・変更が可能である。また、各変形例は、複数又は全部を組み合わせることが可能である。なお、実施例1又は各変形例において説明した処理内容をプログラムとして記録媒体に記録し、このプログラムが記録された記録媒体をコンピュータに読み取らせて前述した処理を情報処理装置に実行させることも可能である。

20

【図面の簡単な説明】

【0105】

【図1】本発明に係るシステムの概略ハードウェア構成の一例を示す図。

【図2】実施例1に係る情報処理装置とクライアントの主要機能構成を示すブロック図。

【図3】分類結果生成手段の詳細な機能構成を示すブロック図。

【図4】記憶手段に記憶される文書データの一例を示す図。

【図5】分類情報の一例を示す図。

【図6】分類情報毎の件数の一例を示す図。

【図7】実施例1における表示画面の一例を示す図。

30

【図8】実施例1における検索処理のフローチャート。

【図9】変形例1に係る情報処理装置とクライアントの主要機能構成を示すブロック図。

【図10】変形例1における表示画面の一例を示す図。

【図11】変形例1における検索処理の一部のフローチャート。

【図12】変形例2における表示画面の一例を示す図。

【図13】変形例3における表示画面の一例を示す図。

【符号の説明】

【0106】

100、900 情報処理装置

101 CPU

40

102 HDD

103 RAM

104 ネットワーク I/F 部

105、106 クライアント

201 検索文字列入力手段

202 分類条件指定手段

203 検索条件生成手段

204、903 検索結果取得手段

205、904 分類情報取得手段

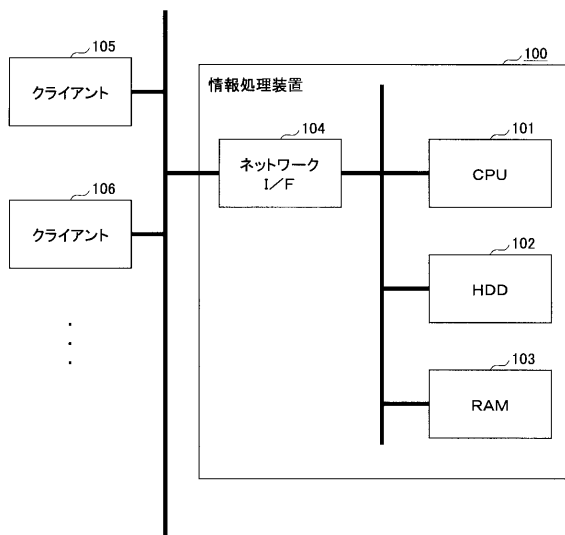
206 検索結果総数取得手段

50

- 207 記憶手段
- 208 分類結果生成手段
- 209 検索結果表示手段
- 210 分類結果表示手段
- 601 概数算出手段
- 602 表示データ生成手段
- 701、702、703 表示領域
- 901 検索結果取得指示手段
- 902 指示受付手段

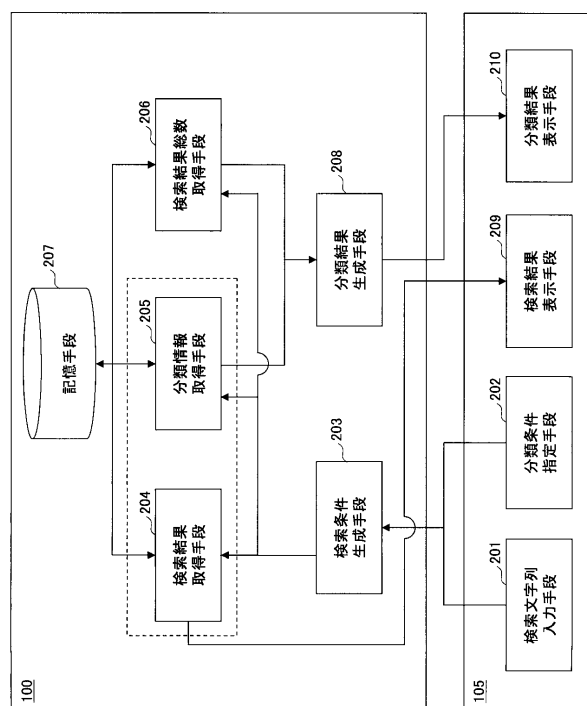
【図1】

本発明に係るシステムの概略ハードウェア構成の一例を示す図



【図2】

実施例1に係る情報処理装置とクライアントの主要機能構成を示すブロック図



【図 3】

記憶手段に記憶される文書データの一例を示す図

ファセット						
ID	タイトル	テキスト	種別	出版年	出版社	...
1	タイトル1	これはサンプル1のテキスト～	書籍	2005	A社	...
2	タイトル2	これはサンプル2のテキスト～	論文	2006	B社	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

分類情報

【図 4】

分類情報の一例を示す図

ID	種別	出版年	出版社	...
1	書籍	2006	A社	...
2	論文	2006	B社	...
5	新聞	2005	B社	...
10	雑誌	2005	A社	...
16	書籍	2004	A社	...

【図 5】

各分類情報の件数の一例を示す図

(A) ファセット:種別

種別	件数
書籍	2
論文	1
新聞	1
雑誌	1

(B) ファセット:出版年

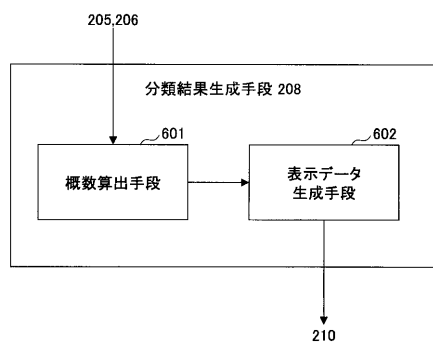
出版年	件数
2006	2
2005	2
2004	1

(C) ファセット:出版社

出版社	件数
A社	3
B社	2

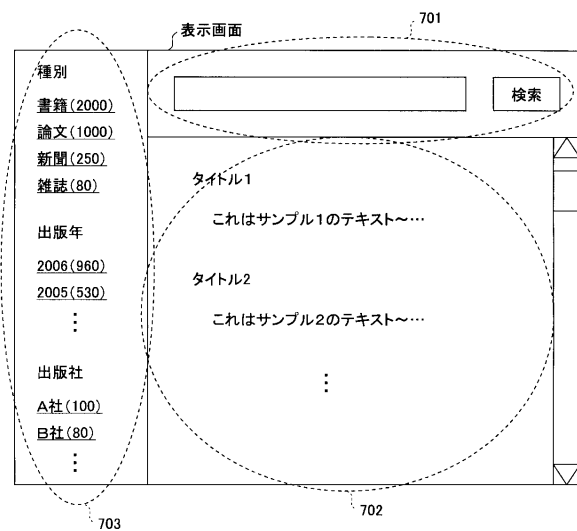
【図 6】

分類結果生成手段の詳細な機能構成を示すブロック図



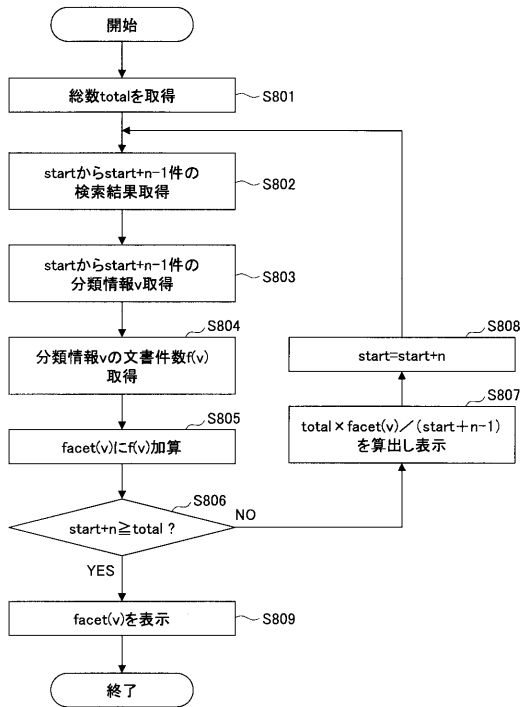
【図 7】

実施例1における表示画面の一例を示す図



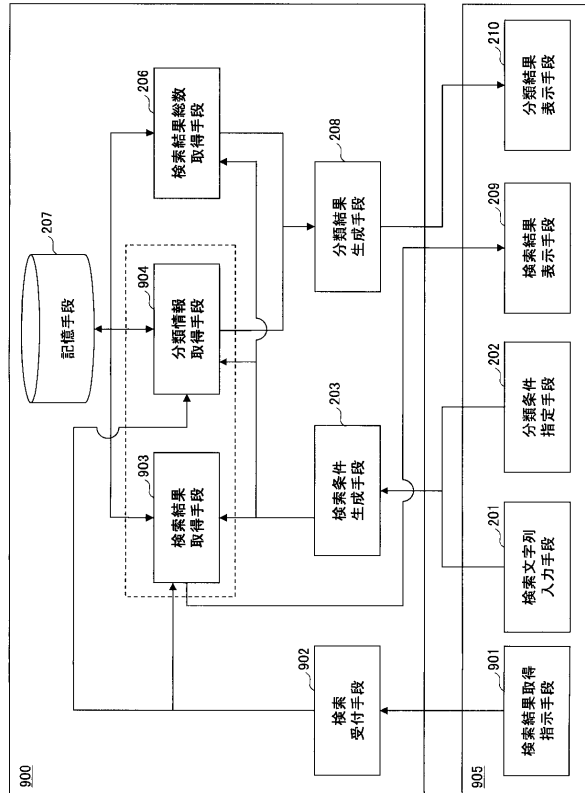
【図 8】

実施例1における検索結果表示処理のフローチャート



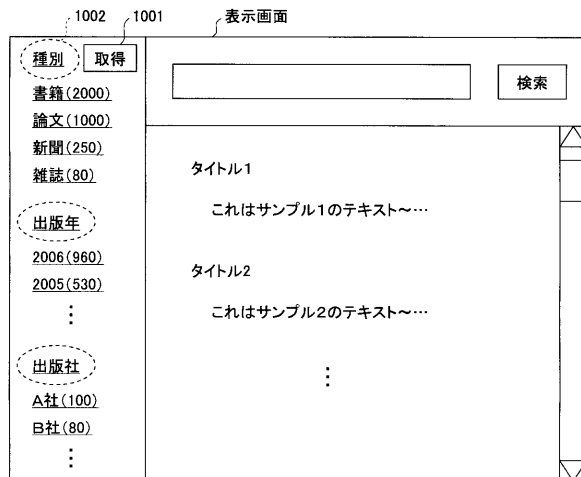
【図 9】

変形例1に係る情報処理装置とクライアントの主要機能構成を示すブロック図



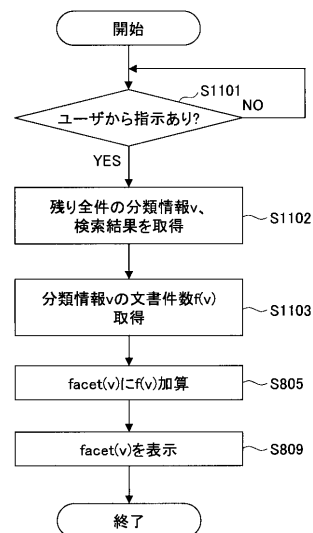
【図 10】

変形例1における表示画面の一例を示す図



【図 11】

変形例1における検索結果表示処理の一部のフローチャート



【図 1 2】

変形例2における表示画面の一例を示す図

1201 表示画面

2000件 / 12345件

種別

書籍(2000)

論文(1000)

新聞(250)

雑誌(80)

出版年

2006(960)

2005(530)

⋮

出版社

A社(100)

B社(80)

⋮

検索

タイトル1

これはサンプル1のテキスト〜…

タイトル2

これはサンプル2のテキスト〜…

⋮

【図 1 3】

変形例3における表示画面の一例を示す図

1301 表示画面

種別

書籍(2000)

論文(1000)

新聞(250)

雑誌(0)

出版年

2006(960)

2005(530)

⋮

出版社

A社(100)

B社(80)

⋮

検索

タイトル1

これはサンプル1のテキスト〜…

タイトル2

これはサンプル2のテキスト〜…

⋮

フロントページの続き

(56)参考文献 特開2005-267597 (JP, A)
特開2002-202983 (JP, A)
特開平11-085764 (JP, A)

(58)調査した分野(Int.Cl., DB名)
G06F 17/30