

【公報種別】特許法第 17 条の 2 の規定による補正の掲載
 【部門区分】第 6 部門第 3 区分
 【発行日】平成 17 年 6 月 30 日 (2005.6.30)

【公開番号】特開 2003-281182(P2003-281182A)
 【公開日】平成 15 年 10 月 3 日 (2003.10.3)
 【出願番号】特願 2002-76923(P2002-76923)
 【国際特許分類第 7 版】

G 0 6 F 17/30

G 0 6 F 17/28

【 F I 】

G 0 6 F 17/30 3 5 0 C

G 0 6 F 17/30 1 7 0 A

G 0 6 F 17/30 2 1 0 D

G 0 6 F 17/28 Z

【手続補正書】

【提出日】平成 16 年 10 月 20 日 (2004.10.20)

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

少なくともテキスト文を含むテキストデータと、当該テキストデータの識別情報とを対応付けるとともに、当該テキスト文に関連した複数の関連情報と、当該複数の関連情報を分類する項目と、当該テキスト文に対応するテキストデータの識別情報とを対応付けるデータベースを検索する情報検索装置において、

前記項目のうち、検索の対象となり得る項目を指定する項目指定情報を記憶する第 1 の記憶手段と、

前記項目指定情報によって指定された項目に分類される関連情報を前記データベースから取得する関連情報取得手段と、

前記関連情報取得手段によって取得された関連情報を、当該関連情報に対応する前記識別情報と対応付けて記憶する第 2 の記憶手段と、

前記項目指定情報によって指定された項目に則した検索条件を取得する検索条件取得手段と、

前記第 2 の記憶手段に記憶された関連情報の中から、前記検索条件に該当する関連情報を特定し、当該関連情報に対応する前記識別情報を特定する検索手段と

を具備することを特徴とする情報検索装置。

【請求項 2】

前記テキストデータからテキスト文を抽出する本文抽出手段と、

前記抽出されたテキスト文を複数の単語に分割して解析する形態素解析手段と、

前記複数の単語の各々が前記テキスト文に出現する回数を計数する出現頻度計数手段とを備え、

前記第 2 の記憶手段は、前記単語と当該単語の出現回数とを、前記テキスト文に対応するテキストデータの識別情報と対応付けて記憶する

ことを特徴とする請求項 1 に記載の情報検索装置。

【請求項 3】

少なくともテキスト文を含むテキストデータと、当該テキストデータの識別情報とを対

応付けるとともに、当該テキスト文に関連した複数の関連情報と、当該複数の関連情報を分類する項目と、当該テキスト文に対応するテキストデータの識別情報とを対応付けるデータベースを検索する情報検索方法において、

前記項目のうち、検索の対象となり得る項目を指定する項目指定情報を記憶装置に記憶する第１の過程と、

前記項目指定情報によって指定された項目に分類される関連情報を前記データベースから取得する第２の過程と、

前記第２の過程において取得された関連情報を、当該関連情報に対応する前記識別情報と対応付けて前記記憶装置に記憶する第３の過程と、

前記項目指定情報によって指定された項目に則した検索条件を取得する第４の過程と、

前記記憶装置に記憶された関連情報の中から、前記検索条件に該当する関連情報を特定し、当該関連情報に対応する前記識別情報を特定する第５の過程と

を具備することを特徴とする情報検索方法。

【請求項４】

少なくともテキスト文を含むテキストデータと、当該テキストデータの識別情報とを対応付けるとともに、当該テキスト文に関連した複数の関連情報と、当該複数の関連情報を分類する項目と、当該テキスト文に対応するテキストデータの識別情報とを対応付けるデータベースを検索するコンピュータを、

前記項目のうち、検索の対象となり得る項目を指定する項目指定情報を記憶する第１の記憶手段、

前記項目指定情報によって指定された項目に分類される関連情報を前記データベースから取得する関連情報取得手段、

前記関連情報取得手段によって取得された関連情報を、当該関連情報に対応する前記識別情報と対応付けて記憶する第２の記憶手段、

前記項目指定情報によって指定された項目に則した検索条件を取得する検索条件取得手段、および、

前記第２の記憶手段に記憶された関連情報の中から、前記検索条件に該当する関連情報を特定し、当該関連情報に対応する前記識別情報を特定する検索手段

として機能させるためのプログラム。

【請求項５】

少なくともテキスト文を含むテキストデータと、当該テキストデータの識別情報とを対応付けるとともに、当該テキスト文に関連した複数の関連情報と、当該複数の関連情報を分類する項目と、当該テキスト文に対応するテキストデータの識別情報とを対応付けるデータベースを検索するコンピュータを、

前記項目のうち、検索の対象となり得る項目を指定する項目指定情報を記憶する第１の記憶手段、

前記項目指定情報によって指定された項目に分類される関連情報を前記データベースから取得する関連情報取得手段、

前記関連情報取得手段によって取得された関連情報を、当該関連情報に対応する前記識別情報と対応付けて記憶する第２の記憶手段、

前記項目指定情報によって指定された項目に則した検索条件を取得する検索条件取得手段、および、

前記第２の記憶手段に記憶された関連情報の中から、前記検索条件に該当する関連情報を特定し、当該関連情報に対応する前記識別情報を特定する検索手段

として機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【手続補正２】

【補正対象書類名】明細書

【補正対象項目名】０００９

【補正方法】変更

【補正の内容】

【 0 0 0 9 】

ここで、上記情報検索装置において、前記テキストデータからテキスト文を抽出する本文抽出手段と、前記抽出されたテキスト文を複数の単語に分割して解析する形態素解析手段と、前記複数の単語の各々が前記テキスト文に出現する回数を計数する出現頻度計数手段とを備え、前記第2の記憶手段は、前記単語と当該単語の出現回数とを、前記テキスト文に対応するテキストデータの識別情報と対応付けて記憶する構成が望ましい。この構成によれば、検索条件として単語が取得された場合に、当該単語を多く含む順にテキストデータの識別情報を特定するといったことが行える。

【 手 続 補 正 3 】

【 補 正 対 象 書 類 名 】 明 細 書

【 補 正 対 象 項 目 名 】 0 0 1 5

【 補 正 方 法 】 変 更

【 補 正 の 内 容 】

【 0 0 1 5 】

取得項目は、グループウェアサーバ20が管理するデータ項目のうち、どの項目を取得するかを指定するものである。詳述すると、グループウェアサーバ20は、文書データに関連する関連情報をデータ項目ごとに分けて記録されたグループウェアファイル22を、文書データごとに備えている。図4は、このグループウェアファイルの一例を示す図である。この図において、文字列「ITEM_NAME」は、データ項目を示すものであり、この文字列「ITEM_NAME」と等号(=)にて結ばれた文字列がデータ項目名を示す。例えば、「ITEM_NAME=Classification」である場合、データ項目名は、「分類(Classification)」となる。また、データ項目名(すなわち、文字列「ITEM_NAME」)の次行がデータ項目名に対応する文書データの関連情報である。具体的には、例えば、文字列「ITEM_NAME=Classification」の次行に記載された文字列「TYPE_TEXT=テクニカルノート」は、データ項目名「分類」に対応する文書データの関連情報が「テクニカルノート」であることを示している。そこで、取得項目は、グループウェアファイル22に含まれるデータ項目名(文字列「ITEM_NAME」によって示されるデータ項目名)のうち、取得すべきデータ項目名を指定する。なお、図示を省略するが、このグループウェアファイル22には、当該グループウェアファイル22が、どの文書データに対応しているかも示されている。

【 手 続 補 正 4 】

【 補 正 対 象 書 類 名 】 明 細 書

【 補 正 対 象 項 目 名 】 0 0 1 6

【 補 正 方 法 】 変 更

【 補 正 の 内 容 】

【 0 0 1 6 】

また、設定ファイル200における重み付け単語は、検索語として頻繁に用いられる単語を指定するためのものである。格納元アドレスは、検索対象となるデータベースが格納されているアドレスを示すものである。詳述すると、グループウェアサーバ20は、上述したように、多数のデータベースを備えるのが一般的であり、このため、どのデータベースを検索対象とするかが特定される必要がある。そこで、アドレスを指定することにより、検索対象となるデータベースを特定するのである。また、格納先アドレスは、上述した格納元アドレスによって特定されるデータベース内の各データから検索用情報に従って抽出した情報を検索用データベース10aに格納するときのアドレスを示すものである。このように、検索対象となるデータベースごとに、異なる格納先アドレスが指定されることで、検索対象となるデータベースごとに抽出した多数の情報を検索用データベース10aに格納することができるようになっている。